

Matematiikan tunteiden mittarin teoreettinen tarkastelu edustavassa suomalaisten nuorten aineistossa

Holm Marja Eliisa
Helsingin yliopisto
Valtiotieteellinen tiedekunta
Tilastotiede
Pro gradu -tutkielma
13.1.2020

Ohjaaja: Kimmo Vehkalahti



HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Tiedekunta/Osasto – Fakultet/Sektion – Faculty

Valtiotieteellinen

Laitos – Institution – Department

Tekijä – Författare – Author

Marja Eliisa Holm

Työn nimi – Arbetets titel – Title

Matematiikan tunteiden mittarin teoreettinen tarkastelu edustavassa suomalaisten nuorten aineistossa

Oppiaine – Läroämne – Subject

Tilastotiede

Työn laji – Arbetets art – Level

Pro gradu -tutkielma /

Kimmo Vehkalahti

Aika – Datum – Month and year

13.1.2020

Sivumäärä – Sidoantal – Number of pages

66 sivua

Tiivistelmä – Referat – Abstract

Tässä tutkimuksessa oppilaiden tunnekokemuksia matematiikassa mitattiin kansainvälisesti käytössä olevalla matematiikan tunnekyselyllä (The Achievement Emotions Questionnaire-Mathematics, AEQ-M; Pekrun, Goetz & Frenzel, 2005). Tämä mittari on oppilaan itsearviointiin perustuva mittari, joka selvittää oppilaiden tunteita matematiikan luokassa, oppimisessa ja testissä. Tämän tutkimuksen tavoitteena on testata mittariin liittyviä teoreettisia malleja, faktorien luotettavuutta, erotteluvaliditeettia ja faktorien erillisyyttä. Lisäksi käsitellään tutkimusotantaa ja puuttuvan tiedon käsittelyä. Tutkimusotos ($n = 1413$) on maantieteellisesti edustava otos 8. luokan oppilaista (14- ja 15-vuotiaat). Tutkimusotos saatiin monivaiheisesti käyttämällä useita otantamenetelmiä kuten kouluihin pohjautuvaa ryväsotantaa, systemaattista otantaa ja implisiittistä osittamista.

Konfirmatorista faktorianalyysia käytettiin tutkimaan teoreettisia AEQ-M-malleja edustavassa tutkimusotoksessa. Tulokset osoittivat, että teoreettinen tunne x tilanne -malli, jossa sekä tunteet että niihin liittyvät tilanteet huomioitiin, kuvasi dataa hyvin. Tässä mallissa kuusi matematiikkaan liittyvää tunnetta (nautinto, ylpeys, viha, ahdistus, häpeä ja tylsyys) mallinnettiin erillisinä latenteina faktoreina, ja eri tilanteiden (luokka, oppiminen ja testi) vaikutus huomioitiin antamalla samassa tilanteessa olevien tunnemuuttujien residuaalien korreloida keskenään. Vaikka tässä mallissa faktoreita vastaavien latenttimuuttujien reliabiliteetit olivat suhteellisen hyviä, tämä tutkimus kuitenkin kyseenalaistaa tunnefaktorien erotteluvaliditeetin ja samaan tilanteeseen liittyvien tunteiden residuaalien väliset korrelaatiot.

Nämä tulokset osoittavat, että mittari on varsin toimiva mittaamaan suomalaisten nuorten tunteita matematiikassa. Lisäksi tutkimustulokset tukevat teoriaa siitä, että akateemisten tunteiden mittaamisessa olisi huomioitava sekä erilliset tunteet että niiden erilaisuus luokka-, oppimis- ja testitilanteessa. Tutkimustulokset herättävät kuitenkin kysymyksen siitä, onko tarkoituksenmukaista yhdistää korkeasti korreloivat tunteet samoiksi faktoreiksi vai käyttää teoreettista mallia korkeista korrelaatioista huolimatta. Vaikka teoria määrittää tunteiden vahvan yhteyden eri tilanteisiin, tämä tutkimus herättää kysymyksen, ovatko kaikki tunteet samassa tilanteessa kuitenkaan merkittävästi yhteydessä. Tulevaisuudessa olisi esitettävä laajempaa analyysia faktorien erillisyydestä ja tunteiden tilannesidonnaisuudesta.

Avainsanat – Nyckelord – Keywords

akateemiset emotiot, erotteluvaliditeetti, faktorianalyysi, matematiikka, konfirmatoriset faktorimallit, otantamenetelmät, puuttuvan tiedon käsittely, nuoret

Sisällys

1. Johdanto	1
2. Tutkimuksen teoreettinen viitekehys	2
2.1 Akateemiset tunteet	2
2.2 Akateeminen tunnekysely matematiikassa (AEQ-M)	4
2.3 Akateemisia emootioita kuvaavat teoreettiset mallit	5
2.3.1 Koko mittaria kuvaavat teoreettiset mallit	5
2.3.2 Reliabiliteetti, erotteluvaliditeetti ja faktorien yhdistäminen	7
3. Analyysimenetelmien teoreettinen viitekehys	9
3.1 Otantamenetelmät	9
3.2 Puuttuva tieto	12
3.3 Odotusarvon maksimointi -menetelmä	13
3.4 Faktorianalyysin määritelmä	17
3.5 Konfirmatorinen faktorianalyysimalli	18
3.6 Esimerkkimalli	21
3.7 Konfirmatorisen faktorianalyysin oletukset	22
3.8 Suurimman uskottavuuden menetelmä	23
3.9 Mallin hyvyttä kuvaavat indeksit	25
3.10 Mittarin reliabiliteetti ja validiteetti	26
4. Tutkimusasetelma ja hypoteesit	30
4.1 Tutkittavat teoreettiset AEQ-M-mallit	30
4.2 Tutkimuskysymykset ja hypoteesit	33
5. Tutkimuksen metodinen toteuttaminen	34
5.1 Tutkimusotos	34
5.2 Tutkimuksessa sovelletun AEQ-M mittarin kuvailu	37
5.3 Tutkimuksen puuttuva data ja sen imputointi	38
5.4 Tutkimuksen kulku	40
6. Tutkimuksen tulokset ja analyysit	41
6.1 Faktorianalyysin ennakko-oletusten tarkastelu	41
6.2 Teoreettisten AEQ-M-mallien ja tutkimushypoteesin testaus	42
6.3 Teoreettisen AEQ-M-mallin reliabiliteetti ja erotteluvaliditeetti	44
6.4 Teoreettisen AEQ-M-mallin parantaminen	45
6.5 Konstruoidun AEQ-M-mallin faktorien erillisyyys	47
7. Pohdinta	51
7.1 Tutkimustulokset suhteessa aikaisempaan teoriaan	51
7.2 Tutkimuksen rajoitukset ja kehitysehdotukset	53
Lähteet	56

1. Johdanto

Oppilaat saattavat kokea moninaisia positiivisia ja negatiivisia tunteita akateemisissa tilanteissa kuten luokkatilanteessa, kotitehtävien teossa tai testitilanteessa (Pekrun, 2006). Oppilaat voivat kokea nautintoa ja ylpeyttä oppiessaan uuden asian luokassa. He saattavat kokea tylsyyttä ja vihaa, kun he joutuvat viettämään useita tunteja matematiikan luokassa. Oppilaat saattavat kokea testiahdistusta ja toivottomuutta, koska he uskovat epäonnistuvansa testissä. Tunteet ovat mukana oppimisprosessissa kokonaisvaltaisesti (Hannula & Holm, 2018; Schutz & Lanehart, 2002).

Koska matematiikan tilanteissa esiintyy useita tunteita, tarvitaan luotettavia mittareita tutkimaan tunteita moninaisesti. AEQ-M-mittari on kehitetty tutkimaan oppilaan moninaisia tunteita kuten nautintoa, ylpeyttä, vihaa, ahdistusta, häpeää, toivottomuutta ja tylsyyttä matematiikan testissä, oppimisessa ja luokassa (Pekrun ym., 2005). Aikaisempi tutkimus on kuitenkin esittänyt ja testannut empiirisesti varsin vähän teoreettisia malleja, jotka kuvaisivat koko AEQ-M-mittarin rakennetta (Moreira, Cunha & Inman, 2018). Lisäksi Moreira ym. (2018) tutkimus kohdistui portugalilaiseen otokseen. AEQ-M-mittarin rakennetta ja luotettavuutta, ei ole testattu suomalaisessa kontekstissa. Siksi tässä tutkimuksessa AEQ-M-mittarin teoreettisia malleja testataan ja tarkastellaan kriittisesti varsin edustavassa suomalaisten nuorten otoksessa.

2. Tutkimuksen teoreettinen viitekehys

Seuraavissa luvuissa (2.1 ja 2.2) tarkastellaan akateemisen tunteiden mittaamiseen liittyvää teoriaa. Luvussa 2.3 esitellään AEQ-M-mittarin teoreettisia malleja, mittarin reliabiliteettia, erotteluväliteettia ja faktorien yhdistämistä. Tutkimushypoteesit perustuvat pääasiassa lukuun 2.3.

2.1 Akateemiset tunteet

Tässä tutkimuksessa tutkitaan tyypillisesti koettuja akateemisia tunteita matematiikassa. Tässä luvussa tarkastellaan tarkemmin siihen liittyvää teoriaa. Akateemiset tunteet (achievement emotions) perustuvat teoriaan, joka voidaan suomentaa ohjausarvoteoriaksi (the control value theory; Pekrun, 2006; Pekrun, Frenzel, Goetz & Perry, 2007). Tässä teoriassa tunteet nähdään joukoksi toisiinsa liittyviä psykologisia prosesseja, jotka sisältävät affektiivisia, kognitiivisia, fysiologisia ja motivaatioon liittyviä komponentteja (Scherer, 2005). Esimerkiksi ahdistukseen voi liittyä hermostuneisuuden kokemusta (affektiivinen), huolestumista (kognitiivinen), hikoilua (fysiologinen) ja halua paeta tilanteesta (motivaatio).

Aikaisemmassa tutkimuksessa on usein keskitytty akateemisiin tuloksiin liittyviin tunteisiin, kuten onnistumiseen ja epäonnistumiseen liittyvään ahdistukseen (Weiner, 1985; Weiner, 2014; Zeidner, 2007). Ohjausarvoteorian mukaan akateemiset tunteet ovat yhteydessä akateemisiin tuloksiin (achievement outcomes emotions), mutta myös akateemisiin aktiviteetteihin (achievement activities emotions; Pekrun, 2006). Akateemisiin aktiviteetteihin liittyviä tunteita ovat nautinto oppimisessa, tylsyys luokkaopetuksessa tai viha ärsyttävän matematiikan tehtävän aikana. Tuloksiin liittyviä emotioita ovat oppilaan kokema ahdistus ennen matematiikan koetta tai oppilaan kokema ylpeys onnistuneen matematiikan tehtävän jälkeen. Tunteet jaetaan tyypillisesti koettuihin tunteisiin tiettyssä tilanteessa (trait achievement emotions) tai hetkellisesti koettuihin tunteisiin tiettyssä tilanteessa tiettyä ajanhetkenä (state achievement emotions; Pekrun, 2006; Pekrun ym., 2007). Oppilailla voi olla hetkellinen

ahdistuksen tunne juuri ennen matematiikan koetta tai heillä voi olla tyypillinen kokemus siitä, että he ovat usein ahdistuneita ennen matematiikan koetta. Lisäksi ohjausarvoteoria määrittää, että akateemiset tunteet eroavat eri tilanteissa, kuten luokassa, opiskelussa ja testissä (Pekrun, Goetz, Frenzel, Barchfeld & Perry, 2011). Esimerkiksi nautinto voi olla erilaista luokassa kuin mitä se on akateemisessa testissä. Siksi akateemisia tunteita pitäisi erotella tilanteiden mukaan.

Tilanteen arvioinnin on määritelty olevan vahvasti yhteydessä tunteiden muodostumiseen (Pekrun, 2006; Scherer, 2005). Ohjausarvoteoria määrittää kaksi keskeistä arviointiprosessia, jotka voivat vaikuttaa tunteiden muodostumiseen. Ensimmäinen prosessi on akateemisten aktiviteettien ja tulosten subjektiivinen hallinta (subjective control). Yksilön arvio siitä, pystyykö hän suoriutumaan akateemisessa tilanteessa, vaikuttaa hänen kokemiinsa tunteisiin (esim. toivottomuus). Myös yksilön arvio siitä, onko hän tai joku toinen aiheuttanut epäonnistumisen tai onnistumisen, vaikuttaa hänen kokemiinsa tunteisiin (esim. häpeä ja viha). Toinen keskeinen arviointiprosessi on yksilön kokema aktiviteettien ja tulosten subjektiivinen arvostus (subjective value). Ohjausarvoteoria määrittää, että tunteita esiintyy vain, jos akateemiset aktiviteetit ja tulokset koetaan riittävän arvokkaiksi.

Seuraavaksi tarkastellaan yksittäisiä tunteita perustuen ohjausarvoteoriaan. Nautinto, tylsyys ja viha liitetään usein akateemisiin aktiviteetteihin (Pekrun, 2006; Pekrun ym., 2007). Oppilas voi kokea nautintoa, jos hän arvostaa matematiikan aktiviteetteja ja kokee, että ne ovat hallittavissa. Tylsyyttä puolestaan esiintyy, jos akateemisista aktiviteeteista puuttuu arvostus. Oppilas voi kokea tylsyyttä, jos matematiikan aktiviteetit ovat merkityksettömiä, yksitoikkoisia, liian vaativia tai liian helppoja. Viha liitetään akateemisiin aktiviteetteihin, mutta myös tuloksiin. Oppilas saattaa kokea vihaa matematiikan aktiviteeteissa, jotka ovat merkityksettömiä, mutta usein hallittavissa ja ratkaistavissa.

Ylpeys, häpeä, ahdistus, toivottomuus ja viha liitetään usein akateemisiin tuloksiin (Pekrun, 2006; Pekrun ym., 2007). Oppilas saattaa kokea ylpeyttä tai häpeää, jos hän kokee itse aiheuttaneensa onnistumisensa tai epäonnistumisensa tärkeässä matematiikan tilanteessa. Oppilas voi puolestaan kokea vihaa, jos hän pitää epäonnistumisen syynä ulkopuolisia

tekijöitä, kuten huonoa opettajaa tai epäselviä tehtäviä. Ahdistus puolestaan liitetään usein epäonnistumiseen, joka koetaan epävarmaksi. Oppilas voi kokea ahdistusta ennen matematiikan koetta tai opiskelua, koska hän on epävarma omasta onnistumisestaan. Toivottomuuden tunne puolestaan liittyy varmaan epäonnistumiseen tai onnistumisen puuttumiseen. Oppilas saattaa kokea toivottomuutta, jos hän on varma epäonnistumisestaan matematiikan kokeessa.

2.2 Akateeminen tunnekysely matematiikassa (AEQ-M)

Tässä tutkimuksessa käytetään matematiikan tunnekyselyä (The Achievement Emotions Questionnaire-Mathematics, AEQ-M; Pekrun ym., 2005) selvittämään oppilaan tunnekokemuksia matematiikan testissä, luokassa ja oppimisessa. Seuraavaksi kuvataan testin rakennetta, jotta tähän mittariin liittyvät teoreettiset mallit olisi helpompi ymmärtää.

AEQ-M on suunniteltu arvioimaan oppilaan akateemisia tunteita suhteessa matematiikkaan. Se pohjautuu siihen näkemykseen, että tunteet on organisoitu oppiainekohtaisesti (Goetz, Frenzel, Pekrun & Hall, 2007). Oppilas voi esimerkiksi kokea tylsyyttä matematiikan tunnilla, mutta ei välttämättä kielten tunnilla. AEQ-M mittaa oppilaiden yleisiä ja tyypillisiä tunnekokemuksia matematiikassa. Muuttamalla tunnemittarin ohjeita sitä voidaan kuitenkin myös käyttää mittaamaan oppilaiden hetkellisiä tunteita matematiikan tilanteessa. Tässä tutkimuksessa mittaria käytettiin mittaamaan nimenomaan tyypillisesti koettuja tunteita matematiikassa. AEQ-M on johdettu akateemisiin tunteisiin liittyvästä yleisestä mittarista (Achievement Emotions Questionnaire, AEQ; Pekrun ym., 2011). AEQ-M koostuu 22 summamuuttujasta, jotka kuvaavat oppilaan seitsemää tunnetta (nautinto, ylpeys, viha, ahdistus, häpeä, toivottomuus ja tylsyys) kolmessa matematiikan tilanteessa: testissä, luokassa ja oppimisessa. Tällä hetkellä AEQ-M on saatavissa ainakin suomeksi, englanniksi, saksaksi ja kiinaksi.

2.3 Akateemisia emootioita kuvaavat teoreettiset mallit

Tässä luvussa tarkastellaan AEQ-M- ja AEQ-mittareista estettyjä teoreettisia malleja sekä akateemisten tunteiden reliabiliteettia ja erotteluvaliditeettia. Tämä luku on yksi tutkielman keskeisimmistä, koska sen sisältö määrittää luvussa 4.2 esitetyt tutkimuskysymykset ja hypoteesit.

Koko mittarin kattavaa mallia on esitetty AEQ-M- (Moreira ym., 2018) tai AEQ-mittareille (Pekrun ym., 2011) vain harvoin. Usein teoreettisissa malleissa on rajoitettu kuvaamaan osaa mittarista, kuten testiin tai luokkaan liittyviä tunteita (Peixoto, Mata, Monteior, Sanches & Pekrun, 2015), kuvaamaan vain osaa tunteista kuten nautintoa, tylsyyttä ja ahdistusta (Lazarides & Buchholzb, 2019; Lichtenfeld, Pekrun, Stupnisky, Reiss & Murayama, 2012) tai esittämään jokainen tunne erillisessä mallissa (Ahmed, van der Werf, Kuyper, & Minnaert, 2013; Frenzel, Thrash, Pekrun, & Goetz, 2007b). Syynä näihin rajoituksiin voi olla, että tunteiden korrelaatiot ovat varsin korkeita (Peixoto ym., 2015; Pekrun ym., 2011), ja siksi voimakkaasti korreloivia tunteita voi olla vaikea esittää samassa mallissa.

2.3.1 Koko mittaria kuvaavat teoreettiset mallit

Aikaisemmassa tutkimuksessa on esitetty varsin harvoin koko AEQ-M-mittaria kattavaa teoreettista mallia (Moreira ym., 2018). Moreira ym. (2018) tutkivat teoreettista AEQ-M-mallia käyttäen konfirmatorista faktorianalyysia ja konstruoivat yhden mallin ohjausarvoteorian (Pekrun, 2006) ja aikaisemman empiirisen tutkimuksen pohjalta (Pekrun ym., 2011). Tämä AEQ-M-malli oli tunne x tilanne -malli, jossa seitsemää matematiikkaan liittyvää tunnetta mallinnettiin seisemänä latenttina faktorina ja eri tilanteiden vaikutus huomioitiin antamalla samassa tilanteessa olevien tunnemuuttujien residuaalien korreloida keskenään. Moreira ym. (2018) havaitsivat, että tunne x tilanne -malli osoitti hyvää mallin hyvyttä.

Pekrun ym. (2011) testasivat aikaisemmin useaa teoreettista mallia yleiselle AEQ-mittarille. AEQ mittaa yhdeksää eri tunnetta (nautinto, toivo, ylpeys, helpotus, viha, ahdistus, toivottomuus, häpeä ja tylsyys) yleisessä akateemisessa tilanteessa. Ensimmäinen malli oli yksifaktorinen tunnemalli, jossa tunteita kuvataan yhdellä bipolaarisfaktorilla (katso kuvio 2: AEQ-M-mittarille vastaava malli). Toinen malli oli kolmifaktorinen tilannemalli, jossa faktorit on eroteltu vain tilanteiden mukaan. Siinä luokkaan, oppimiseen ja testiin liittyvät akateemiset tunteet ovat latautuneet kolmelle tilannefaktorille (katso kuvio 3: AEQ-M-mittarille). Kolmas malli oli yhdeksänfaktorinen tunnemalli, jossa faktorit ovat eroteltu vain erillisten tunteiden mukaan (katso kuvio 4: AEQ-M-mittarille). Neljäs malli oli jo aikaisemmin estetty tunne x tilanne -malli (katso kuvio 5: AEQ-M-mittarille), jossa yhdeksää yleistä akateemista tunnetta mallinnettiin yhdeksänä latenttina faktorina ja eri tilanteiden vaikutus huomioitiin antamalla samassa tilanteessa olevien muuttujien residuaalien korreloida keskenään. Pekrun ym. (2011) havaitsivat, että tämä tunne x tilanne -malli kuvasi dataa kaikkein parhaiten neljästä teoreettisesta mallista.

Aikaisempien tutkimusten mukaan akateemisia tunteita voidaan parhaiten kuvata tunne x tilanne -mallilla, jossa tunteet esitetään erillisinä tunteita ja tunteiden liittyminen eri tilanteisiin huomioidaan (Moreira ym., 2018; Pekrun ym., 2011). Kuitenkin vain harva tutkimus on testannut tätä tunne x tilanne -mallia AEQ-M-mittarille (Moreira ym., 2018). Mittarin luotettavuuden parantamiseksi teoreettisia AEQ-M-malleja olisi testattava eri kulttuurikonteksteissa. Vaikka tämä tunne x tilanne -malli pohjautuu teoreettiseen näkökulmaan siitä, että tunteiden tilannesidonnaisuus olisi huomioitava, on residuaalien välisiä korrelaatioita samaa tilannetta kuvaavien muuttujien välillä suhteellisen paljon ($n > 40$). Akateemisiin tunteisiin liittyviä teoreettisia malleja olisi tarkasteltava kriittisesti ja testattava, onko malleja mahdollista parantaa tarkastelemalla mallissa esitettyjä residuaalien välisiä korrelaatioita.

2.3.2 Reliabiliteetti, erotteluvaliditeetti ja faktorien yhdistäminen

Vaikka aikaisempi tutkimus on harvoin esittänyt teoreettisia AEQ-M-malleja, useat tutkimukset ovat osoittaneet mittarin tunnemuuttujille hyvää reliabiliteettia (Ahmed ym., 2013; Frenzel ym., 2007b; Holm, Hannula & Björn, 2017; Lazarides & Buchholzb, 2019; Moreira ym., 2018; Peixoto ym., 2015; Pekrun ym., 2005). Tutkimukset ovat kuitenkin havainneet varsin korkeita korrelaatioita latenttien tunteiden välillä (Moreira ym., 2018; Peixoto ym., 2015; Pekrun ym., 2011), jolloin mittarin erotteluvaliditeetti voi olla kyseenalainen. Moreira ym. (2018) havaitsivat korkeita korrelaatioita nautinnon ja ylpeyden (.82), vihan ja ahdistuksen (.88), ahdistuksen ja häpeän (.89) sekä vihan ja tylsyyden (.95) välillä tunne x tilanne -mallissa, ja kyseenalaistivat erityisesti vihan ja tylsyyden erotteluvaliditeetin. Peixoto ym. (2015) käyttivät tunteiden tutkimiseen yleisestä AEQ -mittarista johdettua versiota ja tutkivat erikseen matematiikan testiin ja luokkaan liittyviä tunnemalleja. Peixoto ym. (2015) havaitsivat korkeat korrelaatiota esimerkiksi ylpeyden ja nautinnon sekä ahdistuksen ja häpeän välillä ($> .90$) ja päättivät siksi poistaa toisen tunnepareista ennen varsinaista analyysia. Pekrun ym. (2011) havaitsivat myös korkeita korrelaatioita latenttien akateemisten tunteiden välillä (esim. häpeä ja ahdistus: .90) tunne x tilanne -mallissa. Pekrun ym. (2011) ehdottivat, että tunteet ovat toisistaan eroteltavissa huolimatta korkeista korrelaatioista latenttien muuttujien välillä.

Näiden tutkimusten mukaan latentit akateemiset tunteet eivät välttämättä toteuta erotteluvaliditeettia, vaikka reliabiliteetti mittarille on korkea. Tunteiden erotteluvaliditeetista on kuitenkin tehty päätelmiä vain korkeiden korrelaatioiden avulla. Tarvitaan tutkimusta, joka selvittää AEQ-M erotteluvaliditeettia systemaattisten matemaattisten kriteerien perusteella.

Vaikka latenttien tunteiden välillä on esitetty korkeita korrelaatioita, aikaisempi tutkimus ei ole yhdistänyt latentteja tunteita samalle faktorille. Syynä tähän todennäköisesti on vahva teoria (katso luku 2.1; Pekrun, 2006), joka tukee moninaisten ja erillisten tunteiden mallia. Siksi faktorien yhdistäminen saattaisi vaatia uuden teorian muodostamista ja vahvaa empiiristä tutkimusta tunteiden yhdistämisestä. Tarvitaan kuitenkin tutkimusta, jossa systemaattisesti testataan malleja, joissa korkeasti korreloivat latentit tunteet olisi yhdistetty yhdeksi faktoriksi

ja näitä yhdistettyjä malleja verrattaisiin malleihin, joissa nämä tunteet olisi esitetty erillisinä faktoreina. Tämä systemaattinen tarkastelu auttaa ymmärtämään teoreettista näkökulmaa siitä, ovatko tunteet erillisiä vai pitäisikö osa tunteista yhdistää yhdeksi latentiksi tunteeksi.

3. Analyysimenetelmien teoreettinen viitekehys

Seuraavissa luvuissa tarkastellaan tutkimuksen taustalla olevia analyysimenetelmiä kuten otantamenetelmiä (3.1), puuttuvan tiedon käsittelyä (3.2–3.3) ja faktorianalyysimenetelmää (3.4–3.10).

3.1 Otantamenetelmät

Tässä luvussa tutkimuksen taustalla olevaa otantamenetelmää esitellään. Teoreettinen tarkastelu perustuu seuraaviin tilastoanalyysin lähteisiin: Laaksonen, 2013 ja Lehtonen & Pahkinen, 2004.

Varsinainen tavoiteperusjoukko (target population) on se joukko, jota yritetään tutkimuksessa tutkia. Kehikkoperusjoukko (frame population) on puolestaan se joukko, josta tutkimusotos otetaan. Tavoiteperusjoukko ja kehikkoperusjoukko eivät aina kohtaa, eli mukana voi olla ylipeittoa (over coverage) tai alipeittoa (under coverage), jotka aiheuttavat myös tutkimuksessa virhettä (coverage errors). Alipeittoa voi esiintyä, jos kehikkoperusjoukosta puuttuu tavoiteperusjoukossa olevia yksiköitä (esim. syntyneet tai muuttaneet henkilöt tai uudet koulut). Ylipeitto puolestaan tarkoittaa yksiköitä, jotka ovat kehikossa, mutta eivät kuulu tavoiteperusjoukkoon (esim. kuolleet ja pois muuttaneet henkilöt, lakkautetut koulut).

Tutkimusotanta (survey sampling) on mahdollisimman edustava otos perusjoukosta. Se otetaan perusjoukosta erilaisia otanta-asetelmia (sampling design) käyttäen. Otoskehikkoon lisätään usein apumuuttujia (auxiliary variables), joita hyödynnetään itse otannan toteuttamisessa. Niiden avulla voidaan parantaa otoksen arvojen estimointia. Apumuuttujina voi olla tilastolukuja (väestömäärä ikäryhmän mukaan) tai tilastoyksikön koodeja tai arvoja (sukupuoli, alue, sosioekonominen tausta, ikä, koulun koko). Apumuuttujia voidaan saada erilaisista rekistereistä tai tilastoista.

Yksi otantamenetelmä on yksinkertaistettu satunnaisotos ilman takaisinpanoa (simple random sampling without replacement, SRSWOR) tai takaisinpanolla (simple random sampling with replacement, SRSWR). SRSWOR-menetelmässä jokaisen otannan jälkeen otoselementti poistetaan otoskehikosta, kun taas SRSWR-menetelmässä se laitetaan takaisin kehysjoukkoon. Molemmissa menetelmissä tutkimusyksiköitä poimitaan satunnaisesti, mutta niissä ei hyödynnetä apumuuttujia.

Ryväsotanta (cluster sampling) on keskeinen otantamenetelmä tässä tutkimuksessa. Siinä otosta ei oteta suoraan tutkimuspopulaatiota edustavasta kehikosta vaan ryppäistä eli alapopulaatioista (esim. pienalue, yritys tai koulu). Tämä mahdollistaa sen, että otoskehikkoa ei määritellä elementtitasolla vaan tarvitaan ainoastaan klusteritason otoskehikko. Kun esimerkiksi tutkitaan peruskoulun oppilaita, voidaan ensin ottaa otos peruskouluista. Kyseessä on yksiasteinen klusteriotos (one-stage cluster sampling), jos kaikki elementit (esim. peruskoulun oppilaat) otosklustereista otetaan varsinaiseen elementtitason otokseen. Jos näistä valitaan vain osa, esimerkiksi koulun tietyt opetusryhmät (8. luokat), on kyse kaksiaasteisesta klusteriotannasta (two-stage cluster sampling).

Toinen keskeinen otantamenetelmä on systemaattinen otanta (systematic sampling, SYS), jossa otos otetaan tietyn otosvälein ($q = N/n$, missä N = perusjoukon koko ja n = haluttu otoskoko). Tässä menetelmässä valitaan ensin perusjoukon ensimmäinen alkio usein satunnaisesti väliltä $[1, p]$. Tämän jälkeen otetaan joka q :th elementti, kunnes otoskoko n on saavutettu. Systemaattinen otanta voidaan ottaa joko satunnaisesti järjestetystä datasta tai datasta, joka on järjestetty jonkun apumuuttujan, kuten alueen, yrityksen, sukupuolen tai sosioekonomisen taustan mukaan (implisiittinen osittaminen). Tämä järjestäminen on keskeistä, jos sen avulla pystytään paremmin kuvaamaan perusjoukon ja varsinaisen tutkimuspopulaation rakennetta. Esimerkiksi peruskoulujen ryväsotannassa koulut voidaan jakaa alueittain, jolloin saadaan mahdollisimman kattava otos eri alueilta. Tämä menetelmä on lähellä ositettua otantaa (stratified sampling), jossa perusjoukko on jaettu osajoukkoihin eli ryppäisiin apumuuttujan mukaan, ja jokaisesta ryppäistä otetaan oma otoksensa.

Otantaan liittyy myös keskeisiä estimaattoreita eli niitä operaatioita ja kaavoja, joita käytetään määrittämään otannan statistiikkaa. Jokaiselle otosalkiolle voidaan määrittää sisällymistodennäköisyys (inclusion probability = π_k , kun $k \in N$) eli todennäköisyys sille, että alkio sisältyy otokseen. Jos kyseessä on satunnaisotos ilman takaisinpanoa, jokaisen elementin on yhtä todennäköistä tulla valituksi (equal probability sampling).

Tällöin määritetään

$$\pi_k = \frac{n}{N} \quad (3.1.1)$$

kaikille elementeille $k \in N$ (n = otoskoko ja N = perusjoukon koko). Jos jokaiselta otosväliltä otetaan yksi elementti, sisällymistodennäköisyys k :lle elementille on

$$\pi_k = \frac{1}{q} = \frac{n}{N}, \quad (3.1.2)$$

jossa q = poimintaväli. Tällöin sisällymistodennäköisyys on sama kuin yksinkertaisella satunnaisella otoksella ilman takaisinpanoa. Lisäksi jokaiselle otoselementille voidaan määrittää asetelmapaino (design weight) siten, että

$$w_k = \frac{1}{\pi_k}, \quad (3.1.3)$$

kun $k \in s$ (otosjoukko s).

Usein referenssiryhmänä käytetään otosta, joka on poimittu yksinkertaisella satunnaisotoksella takaisinpanolla. Tämän avulla määritetään muiden monimutkaisten otantamenetelmien tehokkuutta.

Yksi tehokkuuden mittari on otanta-asetelmavaikutus eli design-efekti:

$$\text{DEFF} = \frac{V_{P(s)}(\hat{t})}{V_{\text{srswor}}(\hat{t})}, \quad (3.1.4)$$

missä \hat{t} on estimaattori tutkimusmuuttujan summalle koko populaatiossa. DEFF siis määritetään kahden otantamenetelmän design-varianssin suhteena: design-varianssi koskien tutkimuksessa käytettyä otantamenetelmää $[p(s)]$ ja design-varianssi koskien referenssiotantamenetelmää (usein SRSWOR). On huomioitava, että tässä design-variansseja ei ole estimoitu vaan ne ovat laskennallisia arvoja. Usein kuitenkin myös varianssi joudutaan estimoimaan otoksesta (varianssin estimaattori = \hat{V}), koska koko populaation arvoja sen määrittämiseen ei tiedetä. Tällöin estimaattori design-efektille on

$$\text{DEFF} = \frac{\hat{V}_{P(s)}(\hat{t})}{\hat{V}_{\text{srswor}}(\hat{t})}. \quad (3.1.5)$$

Jos $\text{DEFF} = 1$, on otosasetelma yhtä tehokas kuin SRSWOR. Jos $\text{DEFF} < 1$, on otosasetelma tehokkaampi kuin SRSWOR. Jos $\text{DEFF} > 1$, on otosasetelma vähemmän tehokas kuin SRSWOR.

3.2 Puuttuva tieto

Tässä luvussa tarkastelen puuttuvan tiedon käsittelyä. Puuttuvat arvot luokitellaan kolmeen päätyyppiin (Laaksonen, 2013; Scheffer, 2002; Vehkalahti & Everitt, 2019). Jos puuttuvat arvot eivät riipu mistään tutkimusmuuttujista tai havaituista arvoista, arvojen määritetään puuttuvan täysin satunnaisesti (missing completely at random, MCAR). Tällöin puuttuville arvoille ei ole mitään systemaattista selitystä. Jos puuttuvat arvot riippuvat vain havaituista arvoista mutta eivät puuttuvista arvoista, arvojen määritellään puuttuvan satunnaisesti (missing at random, MAR). Jos puuttuvat arvot riippuvat havaituista arvoista mutta myös puuttuvista arvoista, arvot eivät puutu satunnaisesti (not missing at random, NMAR). Koska puuttuvan

tiedon arvoja ei ole tiedossa, NMAR-tapauksen määrittäminen on lähes mahdotonta. Usein ajatellaan, että datan puuttuminen on vähintään satunnaista.

Puuttuvaa dataa voidaan käsitellä usealla tavalla (Laaksonen, 2013; Scheffer, 2002; Vehkalahti & Everitt, 2019). Puuttuva tieto voidaan esimerkiksi poistaa tai paikata mahdollisimman hyvällä imputointimenetelmällä. Imputoinnilla tarkoitetaan aineiston puuttuvien tai käyttökelvottomien arvojen täydentämistä tilastollisin menetelmin. Imputointimenetelmät luokitellaan yleensä kahteen päätyyppiin. Näistä yksinkertaisempi on yksivaiheinen imputointi (single imputation), jossa imputointi tehdään vain kerran. Näitä yksivaiheisia menetelmiä ovat esimerkiksi regressiopaikkaus (regression imputation) ja odotusarvon maksimointi (expectation-maximization, EM). Toinen puuttuvan tiedon korvaamismenetelmä on monivaiheinen imputointi (multiple imputation, MI), jossa imputointi tehdään useaan kertaan. Tällainen menetelmä on esimerkiksi Bayesian monivaiheinen paikkaus.

3.3 Odotusarvon maksimointi -menetelmä

Tässä tutkimuksessa käytettiin EM-algoritmia paikkaamaan puuttuvaa tietoa. EM on lyhenne sanoista expectation-maximization eli odotusarvon maksimointi. Vaikka tämä metodi on yksitasoinen imputointimenetelmä, sen on osoitettu olevan varsin tehokas (Musil, Warner, Yobas & Jones, 2002; Rubin, Witkiewitz, Andre & Reilly, 2007). EM-algoritmi on iteroiva algoritmi uskottavuusfunktion suurimman arvon selvittämiseksi. E viittaa odotusarvoon ja M maksimointiin. EM-estimointi vaihtelee siis odotetun askeleen suorittamisen (expectation step, E-step) ja maksimointiaskeleen välillä (maximization step, M-step). Näitä vaiheita toistetaan vuorotellen, kunnes saavutetaan optimiarvo. Seuraavaksi esitetyt matemaattiset määritelmät EM-algoritmilta perustuvat kahteen tilastolliseen lähteeseen (Dempster, Laird & Rubin, 1977; McLachlan & Krishnan, 2007).

Oletetaan, että y on havaittu data, ja puuttuvaa ei-havaittua dataa kuvataan vektorilla z . Ensimmäisessä E-askeleessa puuttuvan datan arvoja z arvioidaan laskemalla niille ehdolliset odotusarvot tunnettujen arvojen y avulla. Toisessa M-askeleessa selvitetään logaritmisien

uskottavuusfunktion $\log L(\Psi)$ maksimoiva parametrivektori $\Psi = (\Psi_1, \dots, \Psi_d)^T$, suurimman uskottavuuden estimaatti, E-askeleessa saatujen arvojen perusteella. Ensin siis E-askeleessa lasketaan odotusarvot puuttuvalle datalle ja sen jälkeen niitä hyödynnetään täydellisen datan logaritmissen uskottavuusfunktion maksimoinnissa. Tämän jälkeen saatua suurimman uskottavuuden estimaattia käyttämällä lasketaan uudet arvot puuttuvalle datalle. Näitä kahta askelta toistetaan, kunnes parametrien estimaatit eivät enää muutu eri vaiheissa.

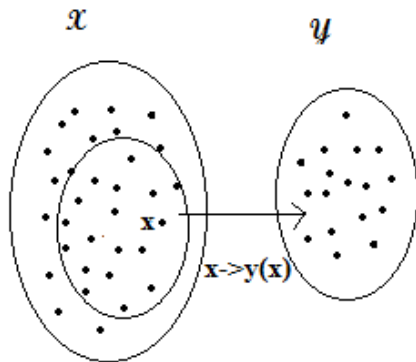
Olkoon Y satunnaisvektori, joka vastaa havaittua dataa y . Luodaan tiheysfunktio $g(y; \Psi)$, jossa $\Psi = (\Psi_1, \dots, \Psi_d)^T$ on tuntemattomien parametrien vektori avaruudesta Ω . Tässä data vektori y on siis epätäydellinen (incomplete data), koska se sisältää puuttuvaa dataa.

Kun puuttuvaa dataa ei ole, datasta käytetään nimitystä täydellinen data (complete data). Oletetaan, että täydellinen data -vektori on x . Merkitään täydelliseen dataa liittyvää funktiota alaindeksillä c . Olkoon nyt $g_c(x, \Psi)$ tiheysfunktio satunnaisvektorille, joka vastaa täydellistä data -vektoria x . Olkoon Ψ edelleen aineiston tuntemattomien parametrien muodostama vektori. Uskottavuusfunktio on tällöin $L_c(\Psi) = g_c(x; \Psi)$. Uskottavuusfunktion maksimi etsitään usein maksimoimalla uskottavuusfunktion logaritmi sen yksinkertaisuuden takia. Tällöin täydellisen datan logaritminen uskottavuusfunktio on

$$\log L_c(\Psi) = \log g_c(x; \Psi). \quad (3.3.1)$$

Tämän avulla saataisiin estimaatti, jos x olisi täysin havaittavissa.

Varsinaisesti EM-algoritmia käytettäessä on olemassa kaksi otosavaruutta \mathcal{X} ja \mathcal{Y} ja useita kuvauksia lähtöavaruudesta \mathcal{X} yhteen malliavaruuden \mathcal{Y} pisteeseen. Sen sijaan, että havaitaan täydellinen datavektori $x \in \mathcal{X}$, havaitaan puutteellinen havaittu datavektori $y = y(x) \in \mathcal{Y}$ (katso kuvio 1).



Kuvio 1. Täydellisen datan \mathcal{X} ja havaitun datan \mathcal{Y} yhteys.

Tästä seuraa, että voidaan olettaa täydellisen datan ja havaitun datan välillä olevan seuraava yhteys

$$g(y; \Psi) = \int_{\mathcal{X}(y)} g_c(x; \Psi) dx, \quad (3.3.2)$$

jossa $\mathcal{X}(y)$ on yhtälön $y = y(x)$ määräämä \mathcal{X} :n osajoukko. Toisin sanoen puutteelliselle datavektorille voidaan määrittää todennäköisin arvo integroimalla täydellisen datavektorin uskottavuusfunktioita.

Koska voidaan havaita vain parametrissa Ψ informaatiota antava mutta puutteellinen aineisto y , uskottavuusyhtälöä ei ole mielekäästä ratkaista suoraan. EM-algoritmin avulla ratkaistaan epätäydelliseen dataan perustuva uskottavuusyhtälö iteroimalla ikään kuin data olisi täydellinen käyttäen sopivasti iteroituja arvoja tuntemattomalle parametrille Ψ .

Ensimmäinen askel voidaan kirjoittaa seuraavasti. Olkoon $\Psi^{(0)}$ jokin alkuarvo tuntemattomille parametreille Ψ . Iterointia varten lasketaan E-askel

$$Q(\Psi; \Psi^{(0)}) = E_{\Psi^{(0)}}\{\log L_c(\Psi) \mid y\}. \quad (3.3.3)$$

Tämä kuvaa logaritmisen uskottavuusfunktion ehdollista odotusarvoa alkuarvolle.

M-askeleessa maksimoidaan funktio $Q(\Psi; \Psi^{(0)})$ Ψ :n suhteen parametrien avaruudessa Ω . Valitaan siis $\Psi^{(1)}$ siten, että jokaiselle Ψ pätee

$$Q(\Psi^{(1)}; \Psi^{(0)}) \geq Q(\Psi; \Psi^{(0)}). \quad (3.3.4)$$

E- ja M- askellusta jatketaan korvaamalla $\Psi^{(0)}$ uudella Ψ :n estimaatilla $\Psi^{(1)}$, jonka avulla ratkaistaan taas uusi estimaatti $\Psi^{(2)}$ maksimoimalla funktiota Q ja niin edelleen.

Tämä EM-algoritmi voidaan kirjoittaa myös yleiseen muotoon seuraavasti:

E-askel. Lasketaan $Q(\Psi; \Psi^{(k)})$, jossa

$$Q(\Psi; \Psi^{(k)}) = E_{\Psi^{(k)}}\{\log L_c(\Psi) \mid y\}, \quad k = 0, 1, 2, \dots \quad (3.3.5)$$

Tätä kaavaa kutsutaan E-askeleeksi EM-algoritmin k :nnella kierrokselle.

K-askele. Valitaan $\Psi^{(k+1)}$ siten, että se maksimoi tämän E-askeleen funktion, jolloin jokaiselle $\Psi \in \Omega$ pätee:

$$Q(\Psi^{(k+1)}; \Psi^{(k)}) \geq Q(\Psi; \Psi^{(k)}). \quad (3.3.6)$$

Tätä maksimointiaskeleta kutsutaan M-askeleeksi. E- ja M-askeleita toistetaan niin kauan, että erotus $L(\Psi^{(k+1)}) - L(\Psi^{(k)})$ muuttuu niin vähän, että iterointi voidaan lopettaa.

3.4 Faktoriansalyysin määritelmä

Faktoriansalyysi on keskeinen menetelmä tarkasteltaessa mittarin rakennetta. Siinä on keskeistä tuntea kaksi muuttujan käsitettä: latentti muuttuja (latent variable) ja havaittu muuttuja (observed or manifest variable; Byrne, 2012; Vehkalahti & Everitt, 2019). Latentti muuttuja on teoreettinen käsite, jota ei voi suoraan havaita ja mitata (esim. viha). Havaitut muuttujat ovat mittarilla mitattuja arvoja (esim. oppilaan raporttoima viha luokassa). Faktorimallissa havaitut muuttujat on asetettu latautumaan latenteille muuttujille eli faktoreille.

Faktoriansalyysi tutkii havaittujen ja latenttien muuttujien välisiä suhteita. Faktoriansalyysissa on kaksi toisistaan poikkeavaa lähestymistapaa: eksploratiivinen faktoriansalyysi ja konfirmatorinen faktoriansalyysi (Byrne, 2012; Vehkalahti & Everitt, 2019). Eksploratiivinen faktoriansalyysi on suunniteltu tilanteeseen, jossa yhteys havaittujen ja latenttien muuttujien välillä on epävarmaa tai ei ole tiedossa. Menetelmässä pyritään määrittämään muuttujien joukosta faktoreita, jotka pystyvät selittämään havaittujen muuttujien vaihtelua ilman tutkijan ennakkotietoja mahdollisesta ilmiöstä. Tällöin menetelmä on aineistolähtöinen. Toisaalta konfirmatorista faktoriansalyysia käytetään, jos tutkijalla on jonkinlainen käsitys latenttien muuttujien rakenteesta ja faktorirakenteesta. Teoriaan ja empiiriseen tutkimukseen perustuen tutkija tekee oletuksia havaittujen arvojen ja niitä edustavien faktoreiden välillä ja tämän jälkeen testaa hypoteesia tilastollisesti. Tässä tutkimuksessa käytetään konfirmatorista

faktorianalyysia tutkimaan AEQ-M-mittarin teoreettisia faktorimalleja. Seuraavaksi tarkastellaan konfirmatorisen faktorianalyysin tilastollista mallia.

3.5 Konfirmatorinen faktorianalyysimalli

Tässä luvussa esitetty faktorianalyysimalli perustuu kolmeen tilastolliseen teokseen: Mustonen, 1995; Rencher & Christensen, 2012 ja Vehkalahti & Everitt, 2019. Konfirmatorinen faktorianalyysimalli on usein määritetty osaksi rakenneyhtälömallinnusta. Molemmissa menetelmissä pyritään testaamaan hypoteettisia malleja faktorien välisistä suhteista. Nämä menetelmät kuitenkin eroavat toisistaan. Rakenneyhtälömallinnuksessa faktorit on usein määritetty ennustamaan toisiaan (polkuanalyysi). Konfirmatorisessa faktorianalyysissä faktorit puolestaan voivat korreloida keskenään, mutta ne eivät ennusta toisiaan.

Ennen varsinaista konfirmatorisen faktorianalyysimallin määrittämistä esitetään yleinen faktorianalyysi malli. Tämä malli voidaan kuvata seuraavasti:

$$y_1 = \mu_1 + \lambda_{11}f_1 + \lambda_{12}f_2 + \dots + \lambda_{1m}f_m + \varepsilon_1$$

$$y_2 = \mu_2 + \lambda_{21}f_1 + \lambda_{22}f_2 + \dots + \lambda_{2m}f_m + \varepsilon_2$$

.

.

.

$$y_p = \mu_p + \lambda_{p1}f_1 + \lambda_{p2}f_2 + \dots + \lambda_{pm}f_m + \varepsilon_p.$$

Tämä malli voidaan kirjoittaa matriisimuodossa seuraavasti:

$$y = \mu + \Lambda f + \varepsilon, \quad (3.5.1)$$

missä

$$\mu = \begin{pmatrix} \mu_1 \\ \cdot \\ \cdot \\ \cdot \\ \mu_p \end{pmatrix}, \Lambda = \begin{bmatrix} \lambda_{11} & \cdots & \lambda_{1m} \\ \vdots & \ddots & \vdots \\ \lambda_{p1} & \cdots & \lambda_{pm} \end{bmatrix}, f = \begin{pmatrix} f_1 \\ \cdot \\ \cdot \\ \cdot \\ f_m \end{pmatrix} \text{ ja } \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_p \end{pmatrix}.$$

Muuttujat $f = (f_1, f_2, \dots, f_m)$ ovat (yhteis)faktoreita eli latentteja muuttujia. Faktoreita on vähemmän kuin alkuperäisiä muuttujia ($m < p$), koska faktorianalyysin ideana on aineiston tiivistäminen. Tässä matriisi Λ on faktorimatriisi, joka muodostuu faktorilatauksista λ_{ij} . Nämä faktorilataukset osoittavat, kuinka jokainen y_i riippuu faktoreista. Faktorilatausta käytetään faktoreiden tulkinnassa. Korkea faktorilataus kuvaa, kuinka paljon muuttujan vaihtelusta voidaan selittää faktorin avulla. Muuttujien vaihtelun ajatellaan jakautuvan kahteen osaan: yhteisvaihteluun, josta olemme kiinnostuneita, ja ominaisvaihteluun, joka voidaan luokitella virhetermiksi tai kohinaksi. Muuttujat $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p)$ ovat ominaisfaktoreita, jotka tulkitaan virhetermeiksi eli residuaaleiksi. Muuttujat $\mu = (\mu_1, \mu_2, \dots, \mu_p)$ ovat havaittujen muuttujien y_i odotusarvoja (keskiarvoja).

Faktoreista koostuvan vektorin $f = (f_1, f_2, \dots, f_m)'$ keskiarvo on 0, ja kovarianssimatriisi on seuraavaa muotoa:

$$\text{cov}(f) = \Phi = \begin{pmatrix} \phi_{11} & \cdots & \phi_{1m} \\ \vdots & \ddots & \vdots \\ \phi_{m1} & \cdots & \phi_{mm} \end{pmatrix}. \quad (3.5.2)$$

Virhevektorin $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p)'$ keskiarvo on 0, ja diagonaalinen kovarianssimatriisi on seuraavaa muotoa:

$$\text{cov}(\varepsilon) = \Psi = \begin{pmatrix} \psi_{11} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \psi_{pp} \end{pmatrix}. \quad (3.5.3)$$

Vektorit f ja ε ovat toisistaan riippumattomia.

On huomiotava, että malli (3.5.1) viittaa tiettyyn kovarianssimatriisin rakenteeseen vektorille y . Kun tämä malli pätee, voidaan populaation kovarianssimatriisi vektorille y määrittää seuraavasti:

$$\Sigma(\theta) = \Lambda\Phi\Lambda' + \Psi. \quad (3.5.4)$$

Tässä Λ , Φ ja Ψ on määritelty kaavojen (3.5.1), (3.5.2) ja (3.5.3) nojalla, ja θ on kokoelma uniikkeja parametreja. Rajoittumattomalle mallille (3.5.1) voidaan määrittää

$$\theta = (\lambda', \phi', \psi').$$

Tässä λ koostuu pm -faktorilatauksesta matriisissa Λ (katso kaava 3.5.1). Parametri ϕ koostuu $m(m + 1)/2$ faktorivarianssista ja kovarianssista symmetrisessä matriisissa Φ (katso kaava 3.5.2). Parametri ψ koostuu p -virhevarianssista diagonaalimatriisissa Ψ (katso kaava 3.5.3). Kun malli (3.5.1) pätee, $\Sigma(\theta)$ on täysin määritelty parametrivektorilla θ .

Konfirmatorisessa faktorianalyysissä pyritään löytämään identifioitavissa oleva malli, jolla on ainutlaatuinen ratkaisu parametrivektorille θ . Malli ei ole identifioitavissa, jos ei ole mahdollista löytää parametreja θ_1 ja θ_2 siten, että $\Sigma(\theta_1) = \Sigma(\theta_2)$. Jos malli ei ole identifioitavissa,

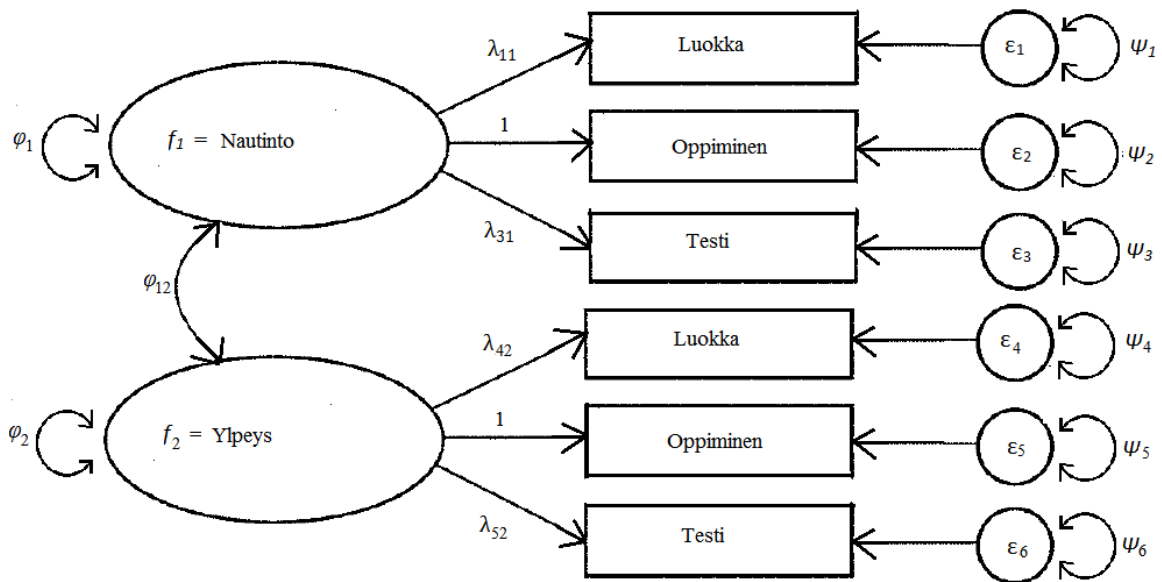
mallille ei ole mahdollista löytää johdonmukaisia estimaatteja. Ei ole kuitenkaan olemassa mitään yleisesti sovittua sääntöä, jolla voitaisiin arvioida mallin identifioituvuutta. Mallin identifiointia voidaan tarkastella järjestysehdon toteutumisen nojalla. Tämä ehto on välttämätön, mutta ei kuitenkaan riittävä ehto mallin identifioitumiselle. Tämä ehto on seuraavaa muotoa:

$$\frac{p(p+1)}{2} \geq q. \quad (3.5.5)$$

Tässä q on estimoitujen parametrien kokonaismäärä ja p on alkuperäisten muuttujien lukumäärä.

3.6 Esimerkkimalli

Seuraavaksi tarkastellaan konfirmatorista faktorimallia yksinkertaisen esimerkin avulla. Havaittuina muuttujina (y) ovat luokkaan, oppimiseen ja testiin liittyvä nautinto ja ylpeys. Kuvio 2 kuvaa kyseistä mallia.



Kuvio 2. Esimerkkimalli.

Malli voidaan kuvata seuraavilla yhtälöillä:

$$\begin{aligned}
 \text{Nautinto luokka} &= \mu_1 + \lambda_{11}f_1 && + \varepsilon_1 \\
 \text{Nautinto oppiminen} &= \mu_2 + f_1 && + \varepsilon_2 \\
 \text{Nautinto testi} &= \mu_3 + \lambda_{31}f_1 && + \varepsilon_3 \\
 \text{Ylpeys luokka} &= \mu_4 && + \lambda_{42}f_2 + \varepsilon_4 \\
 \text{Ylpeys oppiminen} &= \mu_5 && + f_2 + \varepsilon_5 \\
 \text{Ylpeys testi} &= \mu_6 && + \lambda_{52}f_2 + \varepsilon_6.
 \end{aligned}$$

Jos malliin lisätään lisää tunteita kuten viha, ahdistus, häpeä ja tylsyys, mallin esitystapa jatkuu samanlaisena. On huomioitava, että esimerkkimallissa on nyt 13 parametria ($q = 6 + 4 + 3 = 13$) ja kuusi havaittua muuttujaa ($p = 6$). Koska saadaan $[6(6 + 1)] / 2 = 21 \geq 13$, tämä esimerkkimalli toteuttaa välttämättömän ehdon mallin identifioitumiselle (katso kaava 3.5.5).

3.7 Konfirmatorisen faktorianalyysin oletukset

Tässä luvussa esitetään kaksi keskeistä oletusta, jotka on hyvä tarkastaa ennen konfirmatorisen faktorianalyysin toteuttamista.

Oletus 1. Konfirmatorisessa faktorianalyysissä oletetaan usein, että havaittujen muuttujien yhteisjakauma on normaalisti jakautunut eli se noudattaa multinormaalijakaumaa (multivariate normality; Rencher & Christensen, 2012; Vehkalahti, 2014). Havaitulla muuttujalla tarkoitetaan yleensä kyselytutkimuksessa kysymyksiä, joiden arvot muuttuvat vastaajien antamien vastausten mukaan. Oletus ei kuitenkaan usein päde käytännössä, sillä harvan mielipidekysymyksen vastaukset ovat jakautuneet normaalisti. Faktorianalyysi sietää poikkeamia hyvin varsinkin, jos aineisto on laaja. Lisäksi analyysissä on mahdollista valita estimaattori, joka ei vaadi multinormaalisuuden oletuksen toteutumista kuten robusti suurimman uskottavuuden estimaattori (maximum likelihood parameter estimates with standard errors that are robust to non-normality, MLR; Muthén, 2008).

Oletus 2. Lisäksi usein oletetaan, että havaitut muuttujat ovat toisistaan riippumattomia (Vehkalahti, 2014). Kyselytutkimuksessa tämä tarkoittaa, etteivät eri vastaajien antamat vastaukset saisi riippua toisistaan ajan tai paikan suhteen.

3.8 Suurimman uskottavuuden menetelmä

Tässä tutkimuksessa käytetään suurimman uskottavuuden menetelmää (maximum likelihood, ML), kun tutkitaan faktorirakennetta konfirmatorisen faktorianalyysin avulla. Tässä luvussa esitetään tätä menetelmää kolmen teoksen pohjalta: Mustonen, 1995; Rencher & Christensen, 2012 ja Vehkalahti & Everitt, 2019.

Yleisesti ottaen konfirmatorisessa faktorianalyysissä pyritään löytämään parametri θ parametrien avaruudesta Θ siten, että erotus otoksen kovarianssimatriisin (S) ja faktorianalyysimallin kovarianssimatriisin $[\Sigma(\theta)]$; katso kaava 3.5.4] välillä on minimoitu.

Otoksen kovarianssimatriisi havaitulle vektorille y on seuraavaa muotoa:

$$S_n = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})' (y_i - \bar{y}).$$

Oletetaan, että malli (3.5.1) pätee. Tällöin suurimman uskottavuuden funktio on seuraavaa muotoa:

$$L[\Sigma(\theta); S_n] = c \times |\Sigma(\theta)|^{-n/2} \exp\left(-\frac{1}{2} \text{tr}\{nS_n[\Sigma(\theta)]^{-1}\}\right).$$

Tässä c on vakio ja riippumaton parametreista θ ja $\Sigma(\theta)$. Parametrin arvoa θ , joka maksimoi $L[\Sigma(\theta); S_n]$, kutsutaan suurimman uskottavuuden estimaatiksi parametrille θ , ja sitä merkitään $\hat{\theta}_{ML}$.

Uskottavuuden suhde, joka liittyy oletettuun faktorianalyysimalliin, on seuraavaa muotoa:

$$LR(\hat{\theta}_{ML}) = \frac{L[\Sigma(\hat{\theta}_{ML}); S_n]}{L[S_n; S_n]}.$$

Tässä osoittaja on uskottavuus arvioituna Σ :n maksimilla oletetussa mallissa ja nimittäjä on uskottavuus arvioituna Σ :n rajoittumattomalla maksimilla ($\Sigma = S_n$).

Kun yhtälöstä otetaan $-2\log LR(\theta)$, saadaan seuraava yhtälö:

$$l(\theta; S_n) = n(\log |\Sigma(\theta)| + \text{tr}\{S_n[\Sigma(\theta)]^{-1}\} - \log |S_n| - p). \quad (3.8.1)$$

Estimoidut faktorilataukset ja varianssi löydetään minimoimalla funktiota $l(\theta; S_n)$. Tämä onnistuu siten, että $\Sigma(\theta)$ lähenee matriisia S_n . Itse asiassa estimaatti $\hat{\theta}_{ML}$ minimoi funktiota $l(\theta; S_n)$ parametrien avaruudessa Θ .

3.9 Mallin hyvyttä kuvaavat indeksit

Kun mallin parametrit on estimoitu, seuraavassa vaiheessa tulee arvioida, miten hyvin malli sopii havaittuun dataan (Vehkalahti & Everitt, 2019). Erilaisia mallin hyvyysindeksejä käytetään arvioimaan mallin sopivuutta konfirmatorisessa faktorianalyysissä. Tässä luvussa esitetään erilaisia hyvyysindeksejä kuten Khiin neliötesti (the chi-square test), CFI (the comparative fit index), TLI (the Tucker–Lewis index), RMSEA (the root-mean square error of approximation) ja RMR (root mean square residual; Byrne, 2012; Marsh, Hau & Wen, 2004; Vehkalahti & Everitt, 2019).

Eräs testi on Khiin neliötesti (Byrne, 2012; Vehkalahti & Everitt, 2019). Se voidaan määrittää seuraavasti

$$\chi^2 = (N - 1) \min l(\theta; S).$$

Tässä N on otoskoko ja $\min l(\theta; S)$ on minimi arvo suurimman uskottavuuden sovitus funktiolle, joka minimoi matriisien $\Sigma(\theta)$ and S_n erotuksen (katso kaava 3.8.1). Tässä testissä on ongelmana, että hypoteettinen malli hylätään helposti, jos otoskoko kasvaa. Siksi on tärkeää käyttää muita mallin hyvyysindeksejä arvioimaan mallin hyvyttä.

CFI ja TLI voidaan laskea seuraavilla kaavoilla (Byrne, 2012):

$$CFI = 1 - [(\chi_H^2 - df_H) / (\chi_B^2 - df_B)] \text{ ja } (3.9.1)$$

$$TLI = [(\chi_B^2 / df_B) - (\chi_H^2 / df_H)] - [(\chi_B^2 / df_B) - 1]. \quad (3.9.2)$$

Näissä kaavoissa H = oletettu malli ja B = perusmalli (baseline model). Tavallisesti määritetään, että malli on hyväksyttävä, jos CFI ja TLI ovat suurempia tai yhtä suuria kuin .90

(Byrne, 2012; Marsh ym., 2004). RMSEA voidaan puolestaan määrittää seuraavalla kaavalla (Byrne, 2012):

$$\text{RMSEA} = \sqrt{\frac{(\chi^2_H - df_H)/N}{df_H}}. \quad (3.9.3)$$

Indeksi-SRMR on kaikkien standardoitujen residuaalien keskiarvo ja vaihtelee välillä [0, 1]. Mallin sopivuus on hyvä, jos RMSEA ja RMR ovat .05 (.06) tai alle, ja hyväksyttävä, jos ne ovat alle .08 (Byrne, 2012; Marsh ym., 2004).

Malleja verrataan keskenään Satorra–Bentlerin χ^2 -erotustestin avulla (Bryant & Satorra, 2012; Satorra & Bentler, 2010). Mikäli testin tulos on merkittävä, vähemmän rajoitettu (less restrictive) perusmalli (baseline model), jossa on vähemmän vapausasteita (enemmän estimoituja parametreja), sopii malliin paremmin kuin rajoittuneempi vertailumalli (more restrictive, nested, comparison model), jossa on enemmän vapausasteita (vähemmän estimoituja parametreja). Kyseinen erotustesti lasketaan seuraavasti:

$$\chi^2_{SB} = (\chi^2_0 \cdot c_0 - \chi^2_1 \cdot c_1)(df_0 - df_1)/(c_0 \cdot df_0 - c_1 \cdot df_1), \quad (3.9.4)$$

jossa χ^2_0 on χ^2 :n arvo rajoitetulle mallille (nested model); χ^2_1 on χ^2 :n arvo vähemmän rajoitetulle perusmallille (baseline model); c_0 on skaalauskorjausfaktori rajoitetulle mallille; c_1 on skaalauskorjausfaktori vähemmän rajoitetulle perusmallille; df_0 on vapausasteet rajoitetulle mallille ja df_1 on vapausasteet vähemmän rajoitetulle perusmallille.

3.10 Mittarin reliabiliteetti ja validiteetti

Mittauksen luotettavuudesta puhuttaessa erotetaan kaksi peruskäsitettä: validiteetti ja reliabiliteetti (reliability; Litwin, 1995; Vehkalahti, 2014). Reliabiliteetissa on usein kyse

luotettavuudesta ja toistettavuudesta. Mittauksen reliabiliteetti on sitä parempi, mitä vähemmän siihen sisältyy mittausvirhettä. Vaikka mittarin reliabiliteetti olisi hyvä, se ei riitä. Mittari voi nimittäin mitata väärääkin asiaa hyvin johdonmukaisesti. Mittarin on oltava myös validi. Mittarin validiteetilla tarkoitetaan sen pätevyyttä mitata juuri sitä, mitä sen on tarkoitus mitata tarpeeksi kattavasti ja tehokkaasti.

Tutkimuksessa virhe sisältää ainakin kaksi komponenttia: satunnaisvirheen ja mittausvirheen (Litwin, 1995). Satunnaisvirhe on ennustamatonta virhettä, ja sitä esiintyy kaikissa tutkimuksissa. Suuri ja edustava tutkimusotos pienentää satunnaisvirhettä. Mittausvirhe viittaa siihen, kuinka hyvin tai huonosti tietty instrumentti toimii annetussa populaatiossa. Reliabiliteetti ilmaisee sen, miten luotettavasti ja toistettavasti käytetty mittaus- tai tutkimusmenetelmä mittaa haluttua ilmiötä. Tutkimuksen reliaabelius tarkoittaa tutkimuksen kykyä antaa ei-sattumanvaraisia tuloksia. Jos mittari on täysin reliaabeli, siihen eivät vaikuta satunnaisvirheet eivätkä olosuhteet.

Reliabiliteetissa erotetaan useita osatekijöitä, kuten stabiliteetti ja konsistenssi (Litwin, 1995). Stabiliteetissa on kysymys mittarin tai menetelmän pysyvyydestä ajassa. Epästabiliilissa mittarissa näkyvät olosuhteiden ja vastaajan vaikutukset helposti. Uudelleentestaus on usein käytetty menetelmä mittaamaan mittarin stabiiliutta (test-retest reliability). Tätä ei valitettavasti pystytä tässä tutkimuksessa määrittämään, koska mittauskertoja oli vain yksi. Lisäksi monissa tapauksissa tämä reliabiliteetin mittaustapa ei ole toteuttamiskelpoinen, sillä huono reliabiliteettikerroin voidaan usein helpommin selittää ajassa tapahtuneilla todellisilla muutoksilla kuin epästabiiililla mittarilla.

Käytännössä reliabiliteetti liitetään pääasiassa mittarin sisäiseen konsistenssiin (Litwin, 1995). Mittarin sisäinen konsistenssi on indikaattori siitä, kuinka hyvin eri väittämät mittavat samaa asiaa. Paljon käytetty tunnusluku sisäisen konsistenssireliabiliteetin mittaamiseksi on Cronbachin alfa (α , Cronbach, 1951; Litwin, 1995). Cronbachin alfa lasketaan muuttujien välisten keskimääräisten korrelaatioiden ja väittämien lukumäärän perusteella. Cronbachin

alfaa käytetään nimenomaan määrittämään skaalojen (summamuuttujien) reliabiliteetti. Cronbachin alfa voidaan määrittää seuraavasti (Cronbach, 1951):

$$\alpha = \frac{n}{(n-1)} \left(1 - \frac{\sum_i V_i}{V_t} \right). \quad (3.10.1)$$

Tässä n = kysymysten lukumäärä, V_t = testipisteiden varianssi ja V_i = väitepisteen varianssi. Hyväksyttävä alfa on .70 tai enemmän (Nunnally, 1978).

Mittarin konsistenssireliabiliteetti voidaan määrittää myös faktoreille. Tähän liittyvä yhdistelmäreliabiliteetti (composite reliability, CR; Fornell & Larker, 1981; Geldhof, Preacher & Zyphur, 2014) lasketaan seuraavasti:

$$CR = \frac{(\sum_{i=1}^k \lambda_i)^2}{(\sum_{i=1}^k \lambda_i)^2 + \sum_{i=1}^k \theta_i}, \quad (3.10.2)$$

jossa λ_i edustaa väittämän faktorilatausta ja θ_i edustaa väittämän varianssia eli virhevarianssia $(\sum_{i=1}^k 1 - \lambda_i^2)$. Jos $CR \geq .07$ toteutuu kaikille faktoreille, on mittarin sisäinen konsistenssireliabiliteetti hyväksyttävä (Fornell & Larker, 1981; Geldhof ym., 2014).

Myös keskimääräisen selitetyn varianssin arvoa (average variance extracted, AVE; Fornell & Larker, 1981; Zait & Berteau, 2011) käytetään määrittämään mittarin konsistenssireliabiliteettia. Se lasketaan seuraavasti

$$AVE = \frac{\sum_{i=1}^k \lambda_i^2}{\sum_{i=1}^k \lambda_i^2 + \sum_{i=1}^k \theta_i}, \quad (3.10.3)$$

jossa λ_i edustaa väittämän faktorilatausta ja θ_i edustaa väittämän varianssia eli virhevarianssia. Itse asiassa AVE ei riipu niinkään väittämien lukumäärästä, sillä jakajana kaavassa on

väittämien lukumäärä $\sum_{i=1}^k \lambda_i^2 + \sum_{i=1}^k \theta_i$. Jos $AVE < .5$, niin mittausvirheen varianssi (yli 50 %) on suurempi kuin itse rakenteen varianssi, ja reliabiliteetti on kyseenalainen. Jos AVE on .50 tai alle, on mittarin konsistenssireliabiliteetti hyväksyttävä (Fornell & Larker, 1981).

Mittarin validiteettia voidaan määrittää erotteluvaliditeetin mukaan. Erotteluvaliditeetin (discriminant validity) tasoa voidaan arvioida vertaamalla latenttien faktoreiden (A, B) korrelaatioiden neliöjuurta pienempään faktorien AVE-arvoon. Matemaattisesti tämä ehto voidaan kirjoittaa seuraavasti:

$$r_{AB}^2 < \min (AVE_A, AVE_B). \quad (3.10.4)$$

Toisin sanoen, jos jokaisen latentin faktoriparin (A, B) korrelaation neliö (r_{AB}^2) on pienempi kuin niiden pienin AVE-arvo, mittarin erotteluvaliditeetti toteutuu (Fornell & Larker, 1981; Zait & Berteau, 2011).

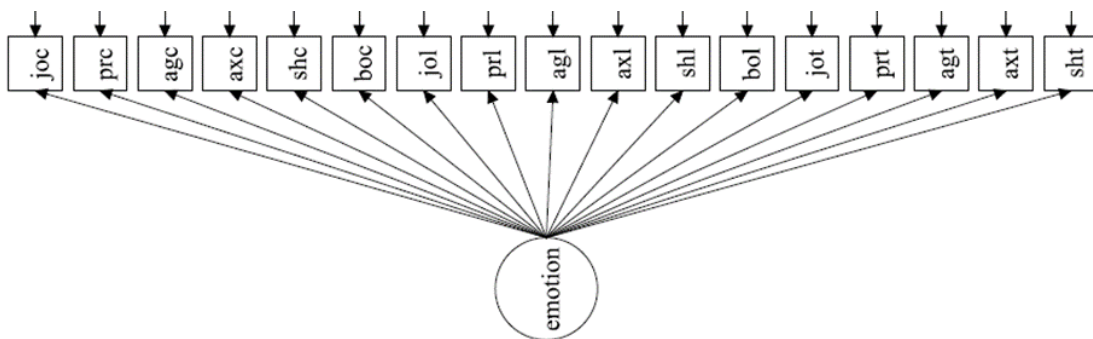
4. Tutkimusasetelma ja hypoteesit

Koska AEQ-M-mittarin teoreettista rakennetta on tutkittu varsin vähän (Moreira ym., 2018), tässä tutkimuksessa testataan aikaisemmassa tutkimuksessa esitettyjä teoreettisia malleja (Moreira ym., 2018; Pekrun ym., 2011). Lisäksi tarkastellaan kriittisesti mittarin erotteluvälikiteettia ja teoreettisten mallien rakennetta. Ennen varsinaista analyysia esitetään tutkimuksen otantamenetelmää.

4.1 Tutkittavat teoreettiset AEQ-M-mallit

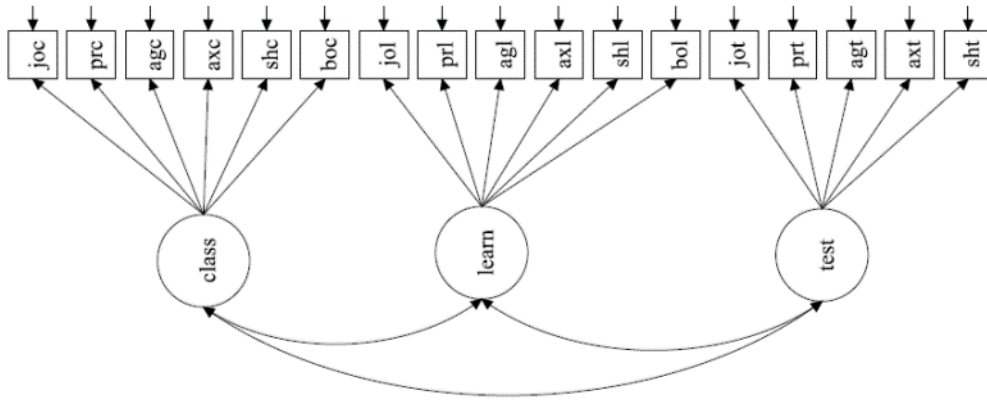
Tässä tutkimuksessa testataan ja tutkitaan neljää teoreettista mallia (katso kuviot 3–5). On huomioitava, että AEQ-M-mittarista ei ole otettu mukaan toivottomuuden tunnetta, koska se edusti vain testiin liittyvää toivottomuutta ja siksi toivottomuuden faktorille olisi tullut vain yksi lataus.

Malli 1. Ensimmäinen teoreettinen malli on esitetty kuviossa 3. Tämä malli on yksifaktorinen tunnemalli, jossa kaikki tunnemuuttujat ovat latautuneet samalle tunnefaktorille. Tämän mallin taustalla on oletus, että tunteita voidaan selittää yhdellä bipolaarisella faktorilla, jossa positiivisilla emotioilla on positiivinen faktorilataus ja negatiivisilla emotioilla on negatiivinen lataus (Pekrun ym., 2011).



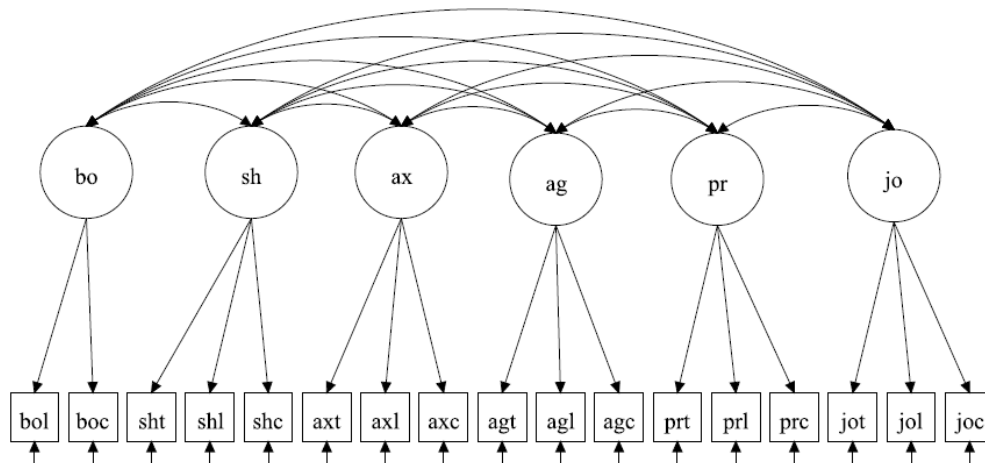
Kuvio 3. Malli 1: yksifaktorinen tunnemalli. jo = nautinto; pr = ylpeys; ag = viha; ax = ahdistus; sh = häpeä; bo = tylsyys; c = luokkatilanne; l = oppiminen; t = testi.

Malli 2. Toinen testattava teoreettinen malli on esitetty kuviossa 4. Tämä malli on kolmifaktorinen tilannemalli, jossa faktorit on eroteltu vain tilanteiden mukaan. Siinä luokkaan, oppimiseen ja testiin liittyvät tunteet ovat latautuneet kolmelle tilanefaktorille.



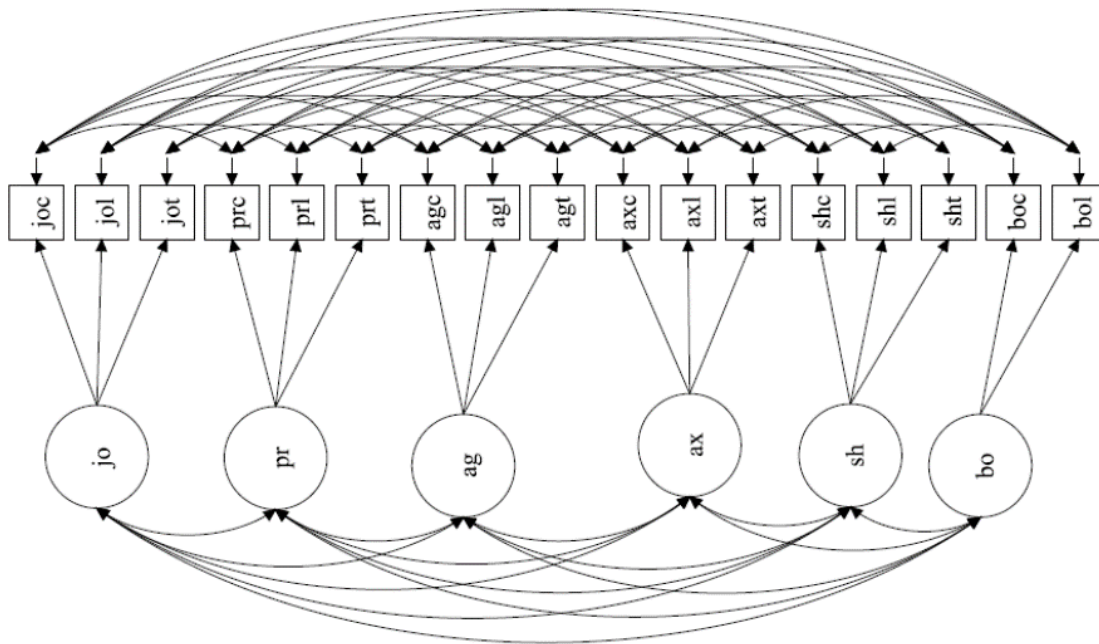
Kuvio 4. Malli 2: kolmifaktorinen tilannemalli. jo = nautinto; pr = ylpeys; ag = viha; ax = ahdistus; sh = häpeä; bo = tylsyys; c = luokkatilanne; l = oppiminen; t = testi.

Malli 3. Kolmas testattava teoreettinen malli on esitetty kuviossa 5. Kolmas malli on kuusifaktorinen tunnemalli, jossa faktorit on eroteltu vain tunteiden mukaan. Se koostuu kuudesta latentista tunnefaktorista (nautinto, ylpeys, viha, ahdistus, häpeä ja tylsyys).



Kuvio 5. Malli 3: kuusifaktorinen tunnemalli. jo = nautinto; pr = ylpeys; ag = viha; ax = ahdistus; sh = häpeä; bo = tylsyys; c = luokkatilanne; l = oppiminen; t = testi.

Malli 4. Neljäs teoreettinen malli on esitetty kuviossa 6. Tämä malli edustaa AEQ-M-mittarin kaksitasoista rakennetta, jossa erilliset tunteet ja niihin liittyvät tilanteet huomioidaan. Malli konstruoidaan käyttämällä korrelaatiolähestymistapaa (Marsh, Byrne & Craven, 1992). Tässä tunne x tilanne -faktorimallissa kuusi matematiikkaan liittyvää tunnetta mallinnetaan latenttina faktorina, ja eri tilanteiden vaikutus huomioidaan antamalla samassa tilanteessa olevien muuttujien residuaalien korreloida keskenään.



Kuvio 6. Malli 4: tunne x tilanne -faktorimalli. jo = nautinto; pr = ylpeys; ag = viha; ax = ahdistus; sh = häpeä; bo = tylsyys; c = luokkatilanne; l = oppiminen; t = testi.

4.2 Tutkimuskysymykset ja hypoteesit

Tässä luvussa esitetään tutkimuskysymykset (1–4) ja asetetaan tutkimushypoteesit (1–2) aikaisemman teorian pohjalta:

Tutkimuskysymys 1. Mikä teoreettisista AEQ-M-malleista 1–4 (katso kuvio 3–6) sopii parhaiten suomalaiseen dataan?

Hypoteesi 1. Ohjausarvoteorian (Pekrun, 2006) ja empiirisen tutkimukseen perusteella (Moreira ym., 2018; Pekrun ym., 2011) tämä tutkimus olettaa, että tunne x tilanne -faktorimalli (malli 4, katso kuvio 6) kuvaasi parhaiten tunteita suomalaisessa aineistossa.

Tutkimuskysymys 2. Millaisia ovat tunnemuuttujien erotteluvaliditeetit ja reliabiliteetit dataa parhaiten kuvaavassa AEQ-M-mallissa?

Hypoteesi 2. Perustuen aikaisempaan tutkimukseen koskien AEQ-M- ja AEQ-mittareita (Ahmed ym., 2013; Frenzel ym., 2007b; Holm ym., 2017; Lazarides & Buchholzb, 2019; Moreira ym., 2018; Peixoto ym., 2015; Pekrun ym., 2005) voidaan olettaa, että teoreettisen mallin erotteluvaliditeetit ovat kyseenalaisia, mutta reliabiliteetit ovat hyviä.

Tutkimuskysymys 3. Koska tunne x tilanne -mallissa on esitetty varsin paljon residuaalien välisiä korrelaatioita samassa tilanteessa olevien tunteiden välillä (Moreira ym., 2018; Pekrun ym., 2011), tässä tutkimuksessa tutkitaan, voiko AEQ-M-mallin rakennetta parantaa residuaalien tarkastelun avulla.

Tutkimuskysymys 4. Onko AEQ-M-mallia mahdollista parantaa yhdistämällä faktoreita, jotka eivät toteuta erotteluvaliditeettia?

5. Tutkimuksen metodinen toteuttaminen

5.1 Tutkimusotos

Tässä luvussa kuvataan tutkimuspopulaatiota ja otantamenetelmiä. Tähän perustuva teoreettinen tarkastelu on esitetty luvussa 3.1.

Tutkimuspopulaationa ovat 8. luokan oppilaat suomalaisista kouluista. Tutkimus toteutettiin kouluissa keväällä 2010. Tutkimuksen kehikkoperusjoukkona käytettiin Tilastokeskuksen vuoden 2008 tilastoa Suomen peruskouluista ($N = 776$), josta oli poistettu kaikki ruotsinkieliset koulut. Tutkimuskehyksessä saattoi olla mukana kouluja, jotka on lakkautettu vuonna 2010 tai uusia kouluja, joita ei löydy enää vuoden 2008 tilastosta. Nämä seikat saattoivat aiheuttaa pientä virhettä (coverage error). Kaikki tutkimusotokseen tulleet koulut olivat kuitenkin toiminnassa myös vuonna 2010.

Tutkimuksen otanta-asetelma koostui useista eri otantamenetelmistä. Tutkimuksessa yhtenä otantamenetelmänä oli ryväs- eli klusteriotanta, jossa otos otettiin varsinaisesti kouluista (klusterit). Itse asiassa kyseessä oli kahden tason klusteriotanta, sillä ensin valittiin otoskoulut ja niistä vain 8. luokan oppilaat. Tässä tutkimuksessa klustereiden kokonaismäärä oli 776.

Tutkimusotoksessa oli myös mukana implisiittistä osittamista, sillä ennen varsinaista otosta koulut oli järjestetty viiteen lääniiin (Länsi-Suomi, Etelä-Suomi, Itä-Suomi, Oulu ja Lappi). Tällöin otos saatiin suhteessa läänissä oleviin kouluihin, ja se vastasi tällöin mahdollisimman hyvin perusjoukkoa. Lisäksi läänissä olevat koulut oli järjestetty aakkosjärjestykseen kaupungin tai kunnan mukaan. Tämä varmisti, että otokseen tuli kouluja monipuolisesti eri kaupungeista tai kunnista. Suurista kaupungeista tuli mukaan useampia kouluja niiden edustavuuden mukaisesti.

Tutkimusotos otettiin tästä järjestetystä datasta systemaattisella otannalla niin, että mukaan tuli 27 koulua, jolloin otantaväli oli $q = 28$, ($q=N/n=776/27=28.74$). Ensimmäinen tutkimuselementti oli listan ensimmäinen koulu. Ensimmäisestä koulusta aloittaminen mahdollisti paremmin perusjoukkoa vastaavien osuuksien saavuttamisen eri lääneistä (huomioi lääni ja aakkosjärjestys). Jokaiselle k :lle koululle voidaan määrittää sisällysmistodennäköisyys systemaattisessa otoksessa kaavan 3.1.2 mukaan:

$$\pi_k = \frac{1}{q} = \frac{n}{N} = \frac{27}{776} \approx 3.5 \text{ \%}.$$

Lopullinen otanta oli edustava otos viidestä Suomen läänistä (Länsi-Suomi: 9 koulua 33 %; Etelä-Suomi: 8 koulua 30 %; Itä-Suomi: 4 koulua, 15 %; Oulu: 3 koulua, 11 % ja Lappi: 3 koulua, 11 %). Koulun koko vaihteli edustavasti pienistä ($n = 65$) isoihin kouluihin ($n = 658$). Tutkimukseen mukaan lähteneistä peruskoulusta poimittiin kaikki 8. luokan oppilaat varsinaisiksi otosoppilaiksi ($n = 1413$).

On huomioitava, että koulujen osuus eri lääneistä vastaa lähes perusjoukon (ilman ruotsinkielisiä kouluja) vastaavia osuuksia lukuun ottamatta Lappia (katso taulukko 1). Syynä tähän on, että ensin eräs Lapin koulu kieltäytyi tutkimuksesta ja seuraava koulu listasta ehdittiin kutsua mukaan otokseen. Kieltäytynyt koulu halusi kuitenkin loppujen lopuksi tulla mukaan tutkimukseen.

Taulukko 1. Koulujen osuudet lääneistä otoksessa ja perusjoukossa.

Läänit	Otos: koulujen osuus (%)	Perusjoukko: koulujen osuus (%)
Länsi-Suomi	33	37
Etelä-Suomi	30	31
Itä-Suomi	15	14
Oulu	11	12
Lappi	11	6
	100	100

Koska koko koulun oppilasmäärät ovat tiedossa sekä perusjoukossa että varsinaisessa tutkimusotoksessa, pystytään näitä arvoja vertaamaan keskenään. Taulukon 2 perusteella otosta voidaan pitää edustavana suhteessa peruskoulun oppilaiden lukumäärään (Otos: $M = 259.95$ ja $SD = 157.77$; Perusjoukko: $M = 221.20$ ja $SD = 170.38$). On huomioitava, että oppilaiden lukumäärän maksimi koko otoksessa ja perusjoukossa eroavat melko paljon. Kuitenkin perusjoukossa on vain 6 peruskoulua, jossa peruskoulun oppilaiden lukumäärä on yli 659.

Taulukko2. Peruskoulun oppilaiden lukumäärät otoksessa ja perusjoukossa.

Estimaatit	Tutkimusotos	Perusjoukko	SRSWOR-otos
Otoskoko (n)	27	776	27
Maksimi (Max)	658	929	573
Keskiarvo (M)	259.96	221.20	229.48
Keskihajonta (s)	157.77	170.38	184.08
Keskivirhe ($S = \frac{s}{\sqrt{n}}$)	30.36	6.12	35.43
Varianssi ($V = s^2$)	24890.58	29029.47	33884.49

Huomio. SRWOR = yksinkertainen satunnaisotos ilman takaisinpanoa.

Seuraavaksi yksinkertainen satunnaisotos ilman takaisinpanoa (SRWOR-otos) otetaan peruskouluista ja määritetään peruskoulun design-efekti koskien peruskoulun oppilaiden lukumäärää. Taulukkoon 2 on kirjattu kyseiset tulokset. Tällöin vastaavaksi design-efektin estimaatiksi saadaan kaavan 3.1.5 mukaan:

$$DEFF = \frac{\hat{V}_{P(s)}(\hat{t})}{\hat{V}_{srswor}(\hat{t})} = \frac{24890.58}{33884.49} = 0.73.$$

$DEFF < 1$ näyttää, että käytetty otanta-asetelma on tehokkaampi kuin yksinkertainen satunnaisotos ilman takaisinpanoa.

5.2 Tutkimuksessa sovelletun AEQ-M mittarin kuvailu

Matematiikkaan liittyviä emootioita arvioitiin Saksassa kehitetyllä AEQ-M-mittarilla (Pekrun ym., 2005). Tätä mittaria on kuvattu tarkemmin luvussa 2.2. AEQ-M on itsearviointiin perustuva instrumentti, jossa oppilaat arvioivat emootioitaan viisiportaisella asteikolla (five-point Likert scale). Mittari koostuu 60 väittämästä, jotka mittaavat seitsemää emootiota luokassa, oppimisessa ja testissä: nautinto, ylpeys, viha, ahdistus, häpeä, toivottomuus ja tylsyys. Aikaisempien teoreettisten mallien mukaisesti (Moreira ym., 2018) tässä tutkimuksessa indikaattoreina ovat eri tilanteisiin liittyvät summamuuttujat, kuten nautinto luokassa, oppimisessa ja testissä ($n = 22$).

Taulukko 3 esittää mittarin tarkempaa rakennetta. Mittarin rakenne sisältää tiettyjä rajoituksia. Ensinäkin toivottomuutta on mitattu vain suhteessa matematiikan testiin ja tylsyyttä on mitattu vain suhteessa matematiikan luokkaan ja opiskeluun. Vaikka muissa tunteissa on huomioitu kolme eri tilannetta, ne sisälsivät usein varsin vähän eri tilanteisiin liittyviä väittämiä. On huomioitava, että tähän tutkimukseen ei sisällytetty toivottomuutta, koska se oli mitattu vain testissä, jolloin faktorille olisi tullut vain yksi lataus.

Englantilainen AEQ-M-mittari käännettiin suomeksi kaksikielisen asiantuntijan toimesta. Se pilotoitiin eräässä suomalaisessa koulussa. Kolmekymmentä 8. luokan oppilasta ja 10 oppilasta, joilla oli erityistarpeita matematiikassa, vastasivat pilottikyselyyn ja antoivat palautetta opettajalle ja tutkijalle. Yhtään negatiivista palautetta ei raportoitu. Oppilaat raportoivat, että he ymmärsivät AEQ-M-kyselyn hyvin ja se oli heistä selkeä toteuttaa. Tässä tutkimuksessa AEQ-M-reliabiliteetti eri tunteiden skaaloille vaihteli hyvästä ($.80 \leq \text{Cronbachin } \alpha < .90$) erinomaiseen ($\alpha \geq .90$; katso kaava 3.10.1).

Taulukko 3. Tunnemittarin rakenne.

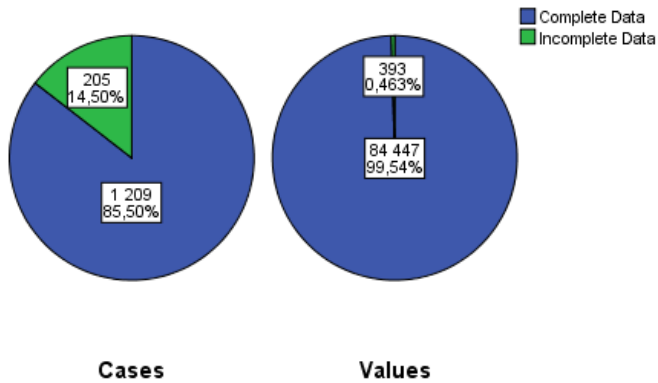
Tunne skaala (lkm)	Luokka		Opiskelu		Testi	
	Koodi	Järjestys kyselyssä	Koodi	Järjestys kyselyssä	Koodi	Järjestys kyselyssä
Nautinto (10)	JO1CB	e1	JO5LD	e22	JO8TD	e46
	JO2CD	e5	JO6LD	e28	JO9TB	e39
	JO3CD	e8	JO7LD	e35	JO10TD	e54
	JO4CD	e12				
Ylpeys (6)	PR1CA	e18	PR3LA	e36	PR5TA	e58
	PR2CA	e19	PR4LD	e32	PR6TA	e59
Viha (9)	AG1CD	e9	AG5LD	e24	AG8TD	e53
	AG2CD	e13	AG6LD	e29	AG9TD	e48
	AG3CD	e15	AG7LD	e34		
	AG4CD	e17				
Ahdistus (15)	AX1CB	e2	AX5LD	e30	AX9TD	e47
	AX2CD	e7	AX6LD	e23	AX10TD	e55
	AX3CB	e3	AX7LD	e26	AX11TB	e38
	AX4CB	e4	AX8LB	e21	AX12TB	e41
					AX13TB	e45
					AX14TB	e43
					AX15TD	e51
Häpeä (8)	SH1CD	e10	SH4LA	e37	SH7TA	e60
	SH2CD	e50	SH5LD	e33	SH8TD	e57
	SH3CD	e16	SH6LD	e27		
Toivottomuus (6)					HL1TB	e40
					HL2TD	e49
					HL3TB	e42
					HL4TB	e44
					HL5TD	e52
					HL6TD	e56
Tylsyys (6)	BO1CD	e11	BO4LB	e20		
	BO2CD	e6	BO5LD	e31		
	BO3CD	e14	BO6LD	e25		

Huomio. JO =nautinto; PR = ylpeys; AG = viha; AX = ahdistus; SH = häpeä; BO = tylsyys; C = luokka; L = oppiminen; T = testi; B = ennen tilannetta; D = tilanteen aikana; A = tilanteen jälkeen.

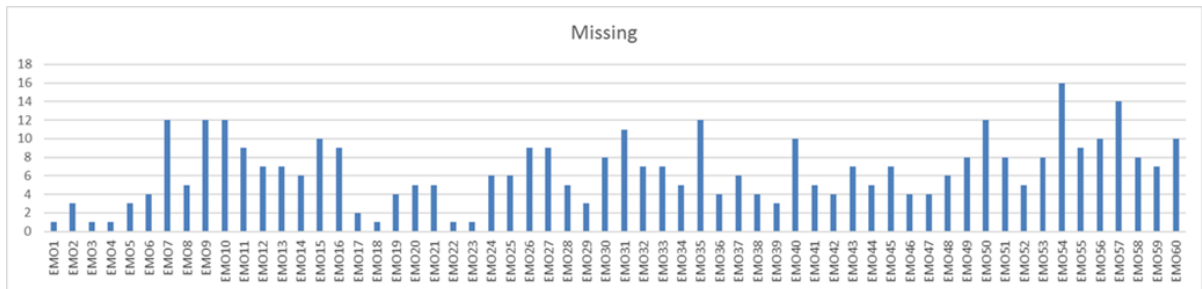
5.3 Tutkimuksen puuttuva data ja sen imputointi

Tässä luvussa kuvataan tutkimuksessa esiintyvää puuttuvaa tietoa ja puuttuvan tiedon imputointiin käytettyä odotusarvon maksimointi -algoritmia (expectation-maximization (EM) algorithm). Puuttuvaan dataan ja EM-algoritmiin liittyvää teoriaa on kuvattu laajemmin luvuissa 3.2 ja 3.3.

Kuten kuvio 7 osoittaa, AEQ-M-väittämissä on vain vähän puuttuvaa tietoa (< 0.5 %). Lisäksi yksittäisissä väittämissä esiintyy usein vain vähän puuttuvaa tietoa (< 1.1 %). Kaikkein eniten puuttuvaa tietoa on väittämissä e54 (1.1 %) ja e57 (.99 %; katso kuvio 8).



Kuvio 7. Puuttuvaa tietoa tapauksissa ja arvoissa.



Kuvio 8. Puuttuvaa tietoa yksittäisissä väittämissä. EMO = emootiokyselyn väittämä.

Seuraavaksi tutkittiin puuttuvan tiedon satunnaisuutta (katso luku 3.2). Tilastollinen testi (Little's test) toteutettiin testaamaan, onko puuttuva tieto täysin satunnaista (MCAR) SPSS25-ohjelmassa. Tässä otoksessa MCAR-hypoteesi ei toteutunut ($\chi^2(5189) = 6186.54, p < .001$), jolloin puuttuva tieto ei ole täysin satunnaista. Kuitenkin voidaan olettaa, että puuttuva tieto on vähintäänkin satunnaista (MAR), eli se ei riipu puuttuvan datan todellisista arvoista. Puuttuva data väittämistä ($n = 60$) imputointiin EM-algoritmeilla käyttäen muita väittämiä apumuuttujina SPSS25-ohjelmalla.

5.4 Tutkimuksen kulku

Tutkimus toteutettiin keväällä 2010. Tutkimusmateriaalit lähetettiin kouluihin postitse. Suomalaisia etnisiä periaatteita humanistis-, sosiaalis- ja käyttäytymistieteissä (National Advisory Board on Research Ethics, 2009) noudatettiin. Lupa tutkimukseen oli pyydetty kunnalta ja rehtorilta, samoin kuin tietoinen suostumus oppilaan vanhemmilta.

Yksityiskohtainen ohje tutkimuksen tekemisestä toimitettiin opettajille. Oppilaille oli annettu 30 minuuttia aikaa toteuttaa AEQ-M matematiikantuntinsa päätteeksi. Opettaja luki ääneen oppilaille kaikki ohjeet ja tarvittaessa AEQ-M-väittämät. Ennen kuin AEQ-M toteutettiin, oppilaille mainittiin heidän vastaustensa luottamuksellisesta käsittelemisestä ja siitä, että ei ole olemassa oikeaa tai väärää vastausta. Heitä pyydettiin vastaamaan oman mielipiteensä mukaan. Kun kaikki mittaukset oli tehty, opettajat keräsivät kaikki materiaalit ja lähettivät ne takaisin tutkijoille postitse.

6. Tutkimuksen tulokset ja analyysit

6.1 Faktorianalyysin ennakko-oletusten tarkastelu

Ennen faktorimallien testaamista tarkastellaan faktorianalyysin ennakko-oletusten 1 ja 2 toteutumista (katso luku 3.7).

Oletus 1. Koska tutkimuksessa käytetään suurimman uskottavuuden estimaattoria (MLR), joka on robusti normaalisuudelle, tutkimusmuuttujien ei tarvitse noudattaa multinormaalijakaumaa. Lisäksi tutkimusotos on kohtuullisen suuri, jolloin parametrien estimointi olisi varsin luotettavaa jopa normaalisuuteen perustuvilla estimaattoreilla.

Oletus 2. Seuraavaksi pohditaan, ovatko tutkimusmuuttujat eli oppilaiden vastaukset AEQ-M-kyselyssä riippumattomia toisistaan ajan ja paikan suhteen. Tämänkaltaista yhteyttä ei pitäisi olla, koska oppilaat ovat täyttäneet kyselyn itsenäisesti, samanaikaisesti ja vain kerran (ei pitkittäistutkimus). Lisäksi aineisto on laaja ja edustava eri kouluista ja alueista (alueellisen riippuvuuden kontrollointi otannassa). Näiden seikkojen takia tämä oletus voidaan olettaa toteutuvaksi.

6.2 Teoreettisten AEQ-M-mallien ja tutkimushypoteesin testaus

Tässä luvussa pyritään vastaamaan luvussa 4.2 esitettyyn tutkimuskysymykseen 1 siitä, mikä teoreettisista AEQ-M-malleista sopii parhaiten suomalaiseen dataan. Kuten luvussa 4.1 on esitetty, testattavat mallit ovat yksifaktorinen malli (malli 1, katso kuvio 3), kolmifaktorinen tilannemalli (malli 2, katso kuvio 4), kuusifaktorinen tunnemalli (malli 3, katso kuvio 5) ja tunne x tilanne -malli (malli 4, katso kuvio 6). Erityisesti testataan hypoteesia 1 siitä, että tunne x tilanne -faktorimalli kuvaaisi parhaiten tunteita suomalaisessa aineistossa.

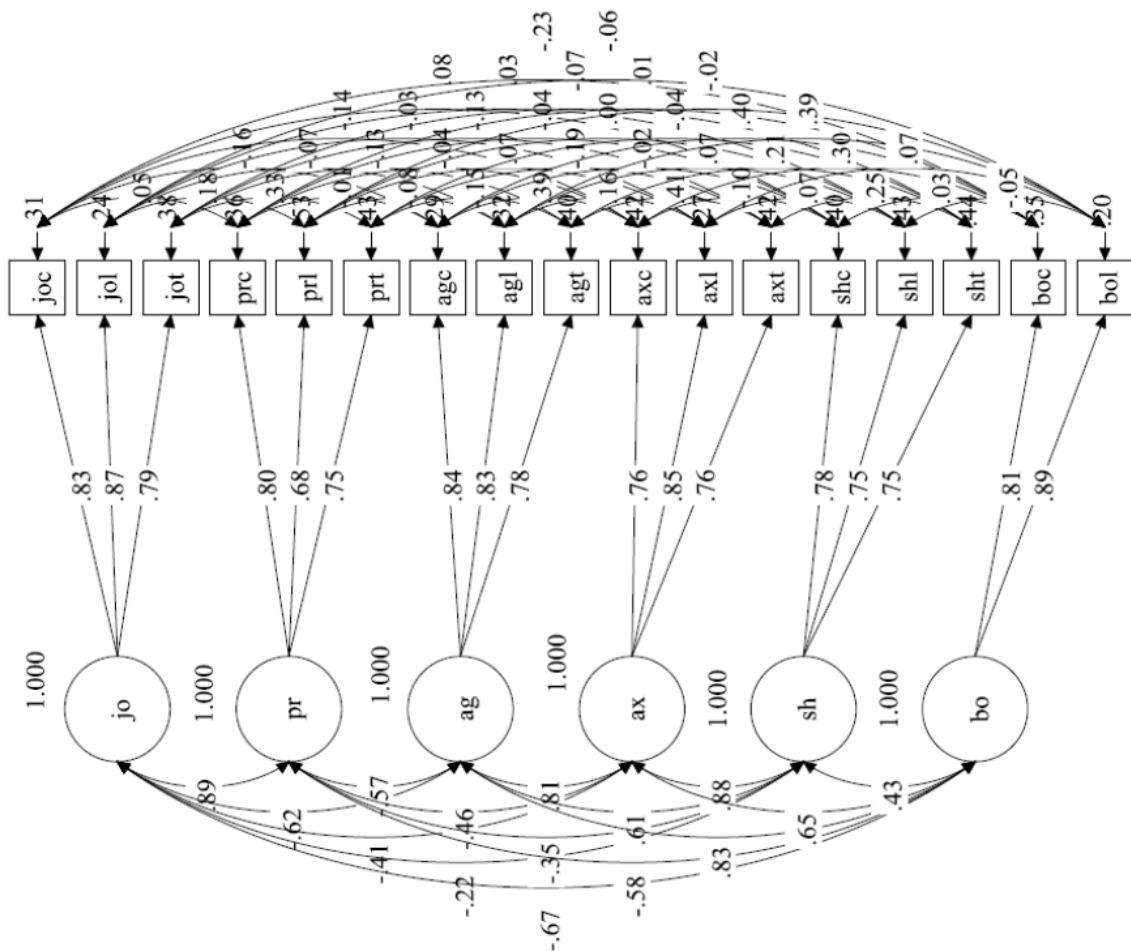
Mallien testaamiseen käytetään konfirmatorista faktorianalyysia, ja se toteutetaan MPLUS-ohjelmassa (versio 7; Muthén & Muthén, 1998–2013). Analyysimenetelmänä on robusti suurimman uskottavuuden menetelmä (MLR). Luvussa 3.9 esitettyjä mallinhyvyysindikaattoreita (CFI, TLI, RMSEA ja SRMR) käytetään arvioimaan sitä, onko mallin hyvyys hyväksyttävä. Lisäksi Satorra–Bentlerin χ^2 -erotustestiä käytetään vertaamaan malleja (katso kaava 3.9.4). Yleisesti on huomioitava, että nämä kaikki testattavat mallit toteuttavat välttämättömän ehdon (3.5.5) mallin identifioitumiselle.

Ensimmäiseksi tulokset osoittavat, että yksifaktorinen tunnemalli kuvaa dataa huonosti [malli 1: $\chi^2(119) = 4876.02, p < .001$, CFI = .61, TLI = .56, RMSEA = .17, SRMR = .13]. Toiseksi myös kolmifaktorinen tilannemalli kuvaa dataa huonosti [malli 2: $\chi^2(116) = 4783.43, p < .001$, CFI = .62, TLI = .55, RMSEA = .17, SRMR = .14], mutta kuitenkin merkittävästi paremmin kuin yksifaktorinen malli Satorra–Bentlerin χ^2 -erotustestin mukaan [$\chi^2(3) = 104.20, p < .001$]. Kolmanneksi kuusifaktorisen tunnemalli osoittaa kohtalaista mallin hyvyttä [malli 3: $\chi^2(104) = 1195.65, p < .001$, CFI = .91, TLI = .88, RMSEA = .09, SRMR = .05] ja merkittävästi parempaa mallin hyvyttä kuin kolmifaktorinen malli Satorra–Bentlerin χ^2 -erotustestin mukaan [$\chi^2(12) = 3369.01, p < .001$].

Lopuksi testattiin tunne x tilanne -mallia, jossa samassa tilanteessa olevien summamuuttujien residuaalien annetaan korreloida keskenään. Analyysi osoitti, että tämä malli kuvaa hyvin dataa [malli 4: $\chi^2(64) = 441.25, p < .001$, CFI = .97, TLI = .94, RMSEA = .06, SRMR = .04] ja osoittaa

merkittävästi parempaa mallin hyvyttä kuin kuusifaktoriainen tunnemalli Satorra–Bentlerin χ^2 -erotustestin mukaan [$\chi^2(40) = 732.59, p < .001$]. Kuvio 9 kuvaa teoreettisen tunne x tilanne -mallin rakennetta, faktorien välisiä korrelaatioita, summamuuttujien latauksia ja residuaalien välisiä korrelaatioita.

Johtopäätöksenä voidaan todeta, että tulokset tukevat hypoteesia 1 siitä, että tunne x tilanne -faktorimalli kuvaa parhaiten tunteita suomalaisessa aineistossa. Lisäksi tulokset tukevat teoriataustaa siitä, että eri akateemisten tunteiden suhdetta voidaan parhaiten kuvata mallilla, joka huomioi erilliset tunteet ja niiden tilannesidonnaisuuden.



Kuvio 9. Teoreettinen tunne x tilanne -malli. jo = nautinto; pr = ylpeys; ag = viha; ax = ahdistus; sh = häpeä; bo = tylsyys; c = luokkatilanne; l = oppiminen; t = testi.

6.3 Teoreettisen AEQ-M-mallin reliabiliteetti ja erotteluvaliditeetti

Edellisessä luvussa osoitettiin, että tunne x tilanne -malli kuvaa parhaiten dataa testatuista AEQ-M-malleista. Seuraavaksi tarkastellaan tämän teoreettisen tunne x tilanne -mallin reliabiliteettia ja erotteluvaliditeettia. Toisin sanoen pyritään testaamaan luvussa 4.2 esitettyä hypoteesia 2 siitä, että teoreettisen mallin erotteluvaliditeetit ovat kyseenalaisia mutta reliabiliteetit ovat hyviä.

Ensimmäiseksi kunkin faktorin pistemäärille lasketaan reliabiliteettikerroin (CR ja AVE) soveltamalla kaavoja 3.10.2 ja 3.10.3. Kuten taulukko 4 esittää, saatujen latenttimuuttujien reliabiliteetit ovat varsin hyviä ja hyväksyttäviä ($CR \geq .70$ ja $AVE > .50$).

Toiseksi tunne x tilanne -mallille määritetään erotteluvaliditeetti. Kuten taulukko 4 esittää, faktorien väliset korrelaatiot ovat varsin korkeita usealle faktoriparille. Itse asiassa joidenkin faktoriparien välinen erotteluvaliditeetti ei ole hyväksyttävä. Toisin sanoen näiden faktoriparien korrelaatioiden neliö ei ole pienempi kuin niiden pienin AVE-arvo (katso kaava 3.10.4). Faktoriparit, jotka eivät erotu riittävästi toisistaan, ovat nautinto ja ylpeys, viha ja ahdistus, ahdistus ja häpeä sekä viha ja tylsyys. Erotteluvaliditeetti tarkastelu on esitetty tarkemmin taulukossa 4.

Taulukko 4. Teoreettisen tunne x tilanne -mallin faktorien väliset korrelaatiot, reliabiliteetit ja erotteluvaliditeetti.

	Nautinto $r(r^2)$	Ylpeys $r(r^2)$	Viha $r(r^2)$	Ahdistus $r(r^2)$	Häpeä $r(r^2)$	CR	AVE
Nautinto						.87	.69
Ylpeys	.886(.78)					.75	.55
Viha	-.623(.39)	-.575(.33)				.86	.67
Ahdistus	-.414(.17)	-.464(.22)	.812(.66)			.83	.63
Häpeä	-.224(.05)	-.352(.12)	.610(.37)	.881(.78)		.80	.58
Tylsyys	-.669(.45)	-.557(.31)	.835(.70)	.649(.42)	.428(.18)	.84	.72

Huomio. r = korrelaatio. Erotteluvaliditeetti ei toteudu, jos $r^2 >$ kuin korreloivien tunteiden pienin AVE (katso kaava 3.10.4).

Johtopäätöksenä voidaan todeta, että tulokset tukevat hypoteesia 2. Toisin sanoen tunne x tilanne -mallin erotteluvaliditeetti on kyseenalainen, vaikka faktoreita vastaavien latenttimuuttujien reliabiliteetit ovat varsin hyviä.

6.4 Teoreettisen AEQ-M-mallin parantaminen

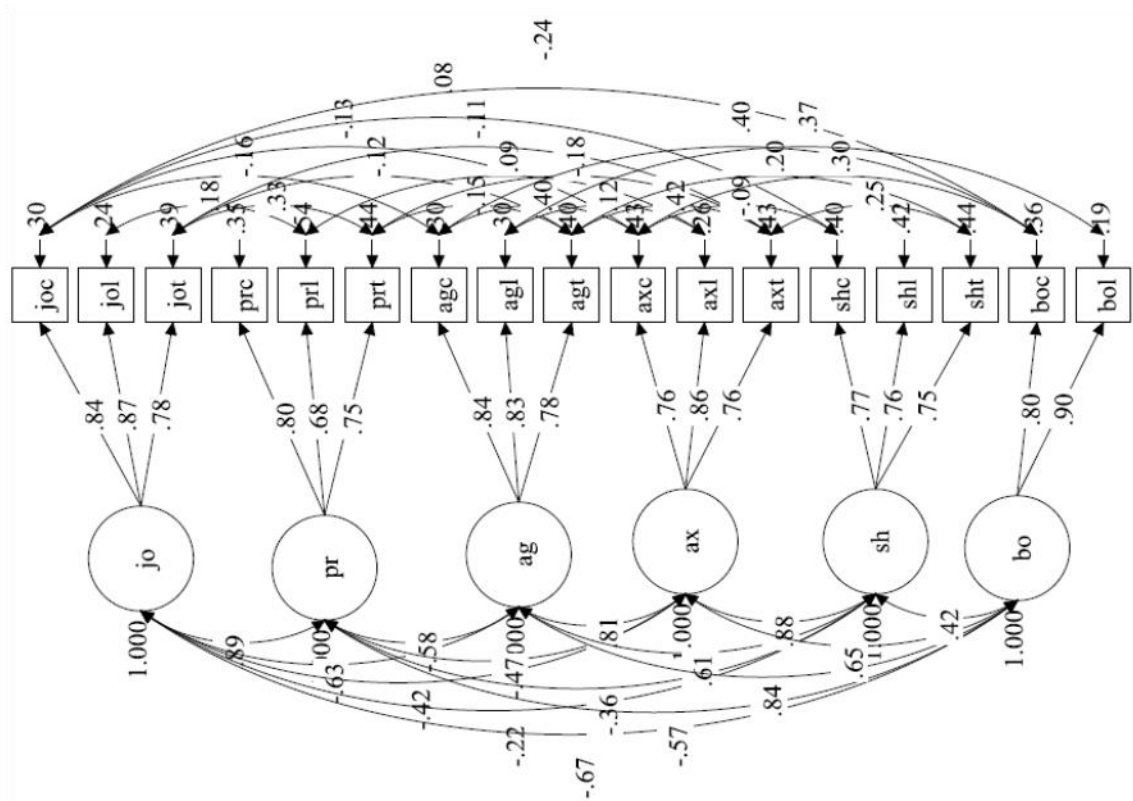
Tässä luvussa vastataan luvussa 4.2 esitettyyn tutkimuskysymykseen 3: Voiko AEQ-M-mallin rakennetta parantaa residuaalien tarkastelun avulla? Luvussa 6.2 havaitsimme, että dataa kuvaa parhaiten teoreettinen tunne x tilanne -malli.

Koska residuaaleja on paljon teoreettisessa tunne x tilanne -mallissa, seuraavaksi tarkastellaan kriittisesti residuaalien välisiä korrelaatioita koskien tätä mallia. Tämä tarkastelu osoittaa, että summamuuttujien residuaalien väliset korrelaatiot eivät aina olleet merkittäviä (50 %, $n = 20$, yhteensä 40 korrelaatiota). Siksi kaikki ei-merkittävät ($p < .05$) korrelaatiot poistetaan yksitellen (PRL with SHL; PRC with AGC; PRL with BOL; PRC with BOC; AGC with SHC; JOL with SHL; PRT with SHT; JOL with BOL; JOL with AXL; PRC with SHC; SHL with BOL; SHC with BOC; JOC with PRC; PRC with AXC; JOL with AGL; AXL with BOL; AXL with SHL; PRL with AGL; JOT with SHT; AGL with SHL). Tätä mallia, jossa ei-merkittävät residuaalien väliset korrelaatiot on poistettu, kutsutaan konstruoiduksi tunne x tilanne -malliksi.

Seuraavaksi tätä konstruoitua tunne x tilanne -mallia (katso kuvio 10) verrataan teoreettiseen tunne x tilanne -malliin, joka sisältää kaikki residuaalien väliset korrelaatiot (katso kuvio 9). Ensinäkin konstruoitu tunne x tilanne -malli osoittaa hyvää mallin hyvyttä [malli 5: $\chi^2(84) = 459.99$, $p < .001$, CFI = .97, TLI = .95, RMSEA = .06, SRMR = .04]. Toiseksi tämä konstruoitu malli (enemmän vapausasteita ja vähemmän estimoituja parametreja) ei myöskään ole merkittävästi huonompi kuin teoreettinen tunne x tilanne -malli [malli 4: $\chi^2(64) = 441.25$, $p < .001$, CFI = .97, TLI = .94, RMSEA = .06, SRMR = .04] Satorra–Bentlerin χ^2 -erotustestin mukaan [$\chi^2(20) = 25.19$, $p = .194$]. Tämän analyysin perusteella konstruoitu tunne x tilanne -malli valitaan kuvamaan tunteita teoreettisen mallin sijasta. Kuvio 10 kuvaa konstruoidun

tunne x tilanne -mallin rakennetta, faktorien välisiä korrelaatioita, summamuuttujien latauksia ja residuaalien välisiä korrelaatioita.

Johtopäätöksenä voidaan todeta, että teoreettista AEQ-M-mallia voidaan parantaa vähentämällä ei-merkittäviä residuaalien välisiä korrelaatioita samaan tilanteeseen liittyvien tunteiden välillä. Tämä tulos myös herättää kriittisen kysymyksen siitä, liittyvätkö kaikki samassa tilanteessa esiintyvät tunteet aina toisiinsa.



Kuvio 10. Konstruoitu tunne x tilanne -malli. Samassa tilanteessa olevien summamuuttujien residuaalien annetaan korreloida keskenään, jos korrelaatio oli merkittävä. jo = nautinto; pr = ylpeys; ag = viha; ax = ahdistus; sh = häpeä; bo = tylsyys; c = luokkatilanne; l = oppiminen; t = testi.

6.5 Konstruoidun AEQ-M-mallin faktorien erillisuus

Lopuksi tarkastellaan vielä konstruoidun tunne x tilanne -mallin reliabiliteettia, erotteluvaliditeettia ja faktorien erillisyyttä. Erityisesti vastataan luvussa 4.2 esitettyyn tutkimuskysymykseen 4 siitä, onko AEQ-M-mallia mahdollista parantaa yhdistämällä faktoreita, jotka eivät toteuta erotteluvaliditeettia.

Ensimmäiseksi konstruoidun mallin faktorien pistemäärille laskettiin reliabiliteettikerroin (CR ja AVE) soveltamalla kaavoja 3.10.2 ja 3.10.3. Kuten taulukko 5 esittää, faktoreita vastaavien latenttimuuttujien reliabiliteetit ovat varsin hyviä ja hyväksyttäviä ($CR \geq .70$ ja $AVE > .50$).

Toiseksi määritetään mallin erotteluvaliditeetti (kaava 3.10.4). Kuten taulukko 5 esittää, faktorien välinen erotteluvaliditeetti ei ole vielääkään hyväksyttävä. Toisin sanoen joidenkin faktoriparien korrelaatioiden neliö ei ole pienempi kuin niiden pienin AVE-arvo. Faktoriparit, joille tämä ehto ei toteudu, ovat edelleen nautinto ja ylpeys, viha ja ahdistus, ahdistus ja häpeä sekä viha ja tylsyys.

Taulukko 5. Konstruoidun tunne x tilanne -mallin faktorien väliset korrelaatiot, reliabiliteetit ja erotteluvaliditeetti.

	Nautinto $r(r^2)$	Ylpeys $r(r^2)$	Viha $r(r^2)$	Ahdistus $r(r^2)$	Häpeä $r(r^2)$	CR	AVE
Nautinto						.87	.69
Ylpeys	.889(.79)					.75	.55
Viha	-.628(.39)	-.582(.34)				.86	.67
Ahdistus	-.417(.17)	-.469(.22)	.809(.65)			.83	.63
Häpeä	-.223(.05)	-.356(.13)	.606(.38)	.880(.77)		.81	.58
Tylsyys	-.671(.45)	-.574(.33)	.838(.70)	.654(.43)	.420(.18)	.84	.72

Huomio. r = korrelaatio. Erotteluvaliditeetti ei toteudu, jos $r^2 >$ kuin korreloivien tunteiden pienin AVE (katso kaava 3.10.4).

Koska mittarin erotteluvaliditeetti on kyseenalainen, seuraavaksi näitä faktoripareja, jotka eivät toteuta erotteluvaliditeettia, yhdistetään yhdeksi faktoriksi. Saadut mallit nimetään yhdistysmalleiksi. Näitä yhdistysmalleja verrataan alkuperäiseen konstruoituun kuuden

faktorin tunne x tilanne -malliin. Kuten taulukko 6 osoittaa, on konstruoitu kuuden faktorin tunne x tilanne -mallin hyvyys kuitenkin parempi kuin näiden yhdistysmallien hyvyys. Toisin sanoen Satorra–Bentlerin χ^2 -erotustestin mukaan alkuperäisen mallin hyvyys on aina merkittävästi parempi kuin mallien, jossa faktoripareja on yhdistetty samalla faktorille.

Taulukko 6. Faktorien yhdistäminen malleissa.

Faktorit	χ^2	<i>df</i>	TLI	CFI	RMSEA	SRMR	<i>c</i>		χ^2_{SB}
1. Malli 5	459.99***	84	.969	.950	.056	.040	1.25		
2. JO_PR	611.86***	89	.957	.935	.064	.047	1.26	1 vs. 2	129.23***
3. SH_AX	669.25***	89	.952	.927	.068	.055	1.25	1 vs. 3	187.86***
4. BO_AG	765.15***	89	.945	.915	.073	.048	1.25	1 vs. 4	305.16***
5. AG_AX	1047.35***	89	.922	.880	.087	.087	1.25	1 vs. 5	587.36***
6. JO_PR; SH_AX; BO_AG	1098.46***	96	.918	.890	.084	.064	1.26	1 vs. 6	608.32***
7. JO_PR; SH_AX_BO_AG	2017.23***	98	.843	.782	.118	.098	1.23	1 vs. 7	1717.30***

Huomio. Malli 5 = konstruoitu tunne x tilanne -malli. JO = nautinto; PR = ylpeys; SH = häpeä; AX = ahdistus; BO = tylsyys; AG = viha. Alaviiva tarkoittaa, että kyseiset faktoriparit on yhdistetty. * $p < .05$; ** $p < .01$; *** $p < .001$.

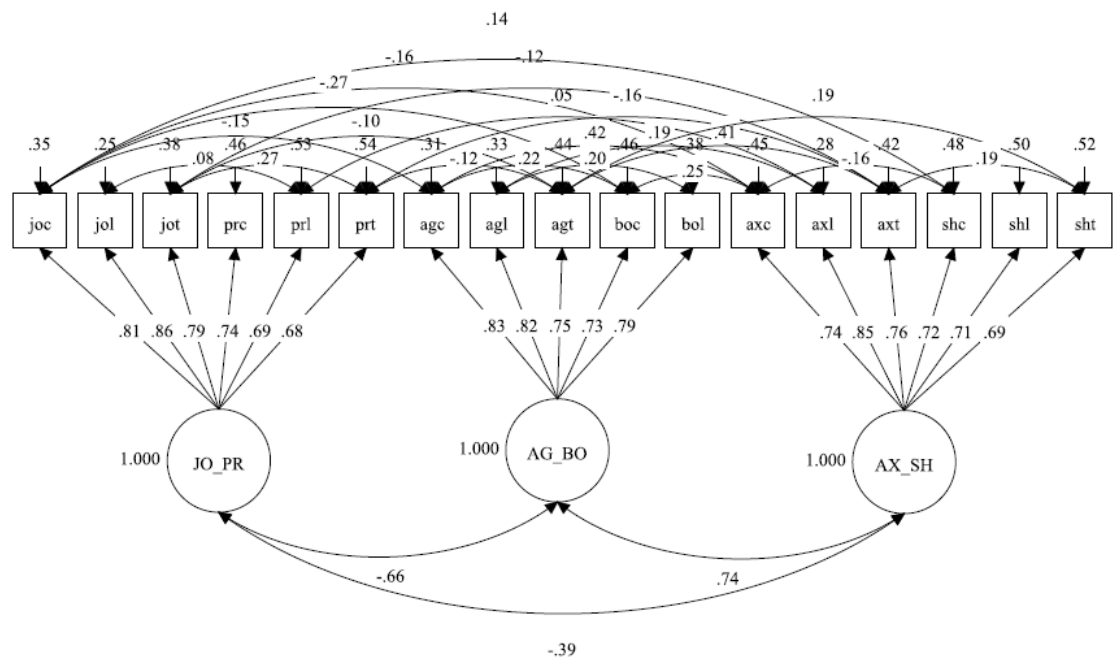
Seuraavaksi tutkittiin tarkemmin kolmifaktorista yhdistysmallia (katso malli 6 taulukossa 6). Koska tässä mallissa latenttien muuttujien korrelaatiot olivat suhteellisen alaiset (–.67, –.39, ja –.74), halutaan selvittää, olisiko tällä yhdistysmallilla riittävää erotteluvaliditeettia. Vaikka tämän kolmifaktorisen mallin hyvyys huononi merkittävästi faktoreita yhdistäessä, oli se kuitenkin kohtalainen. On huomioitava, että kaksifaktorisen yhdistysmallin hyvyysindeksit ovat jo varsin heikkoja (katso malli 7 taulukossa 6).

Taulukko 7. Kolmifaktorisen yhdistysmallin faktorien väliset korrelaatiot, reliabiliteetit ja erotteluvaliditeetti.

	JO_PR $r(r^2)$	AG_BO $r(r^2)$	CR	AVE
JO_PR_			.89	.58
AG_BO	–.659(.43)		.88	.62
AX_SH	–.394(.16)	.738(.54)	.86	.56

Huomio. r = korrelaatio; JO = nautinto; PR = ylpeys; AG = viha; BO = tylsyys; AX = ahdistus; SH = häpeä. Erotteluvaliditeetti ei toteudu, jos $r^2 >$ kuin korreloivien tunteiden pienin AVE (katso kaava 3.7.4).

Nyt myös tälle kolmifaktoriselle mallille laskettiin reliabiliteettikerroin (CR ja AVE) soveltamalla kaavoja 3.10.2 ja 3.10.3. Kuten taulukko 7 esittää, faktoreita vastaavien latenttimuuttujien reliabiliteetit ovat varsin hyviä ja hyväksyttäviä ($CR \geq .70$ ja $AVE > .50$). Lisäksi määritetään mallin erotteluvaliditeetti (kaava 3.10.4). Kuten taulukko 7 esittää, faktorien välinen erotteluvaliditeetti on nyt hyväksyttävä. Toisin sanoen faktoriparien korrelaatioiden neliö on pienempi kuin niiden pienin AVE-arvo. Kuvio 11 kuvaa konstruoidun kolmifaktorisen yhdistysmallin rakennetta, faktorien välisiä korrelaatioita, summamuuttujien latauksia ja residuaalien välisiä korrelaatioita.



Kuvio 11. Kolmifaktorinen yhdistysmalli. jo = nautinto; pr = ylpeys; ag = viha; ax = ahdistus; sh = häpeä; bo = tylsyys; c = luokkatilanne; l = oppiminen; t = testi.

Johtopäätöksenä voidaan todeta, että mallin hyvyysindeksien perusteella tunteita pitäisi tarkastella erillisinä tunteina huolimatta siitä, että erotteluvaliditeetti on kyseenalainen. Toisin sanoen tulokset osoittavat, että AEQ-M-mallin hyvyttä ei ole mahdollista parantaa yhdistämällä faktoreita, jotka ei toteuta erotteluvaliditeettia. Toisaalta tulokset osoittavat, että mittarin erotteluvaliditeetista tulee hyväksyttävä, kun korkeasti korreloivat faktoriparit nautinto ja ylpeys, viha ja tylsyys sekä ahdistus ja häpeä yhdistetään samalle faktorille. Tällöin tulevaisuudessa olisi mietittävä, onko valittava AEQ-M-malli, jolla on hyvä erotteluvaliditeetti

(katso kuvio 11) vai AEQ-M-malli, jolla on hyvä mallin hyvyys mutta huono erotteluvaliditeetti (katso kuvio 10).

7. Pohdinta

Tämän tutkimuksen päätavoitteena on tutkia kriittisesti teoreettisia AEQ-M-malleja laajassa suomalaisessa nuorten aineistossa. Tämä tutkimus on merkityksellinen, koska koko AEQ-M-mittaria kattavaa mallia ei ole esitetty suomalaisessa kontekstissa. Tutkimus havaitsi teorian ja empiirisen tutkimuksen mukaisesti (Moreira ym., 2018; Pekrun, 2006), että teoreettinen tunne x tilanne -faktorimalli (katso luku 4.1, kuvio 6) kuvasi parhaiten tunteita suomalaisessa aineistossa. Tämä tutkimus osoittaa, että AEQ-M-mittaria voidaan käyttää varsin luotettavasti mittaamaan suomalaisten tunteita matematiikassa. Tutkimus nostaa kuitenkin esiin muutamia kriittisiä seikkoja, jotka liittyvät mittarin erotteluväliditeettiin, residuaalien välisiin korrelaatioihin ja latenttien tunnefaktoreiden yhdistämiseen. Seuraavassa luvussa tuloksia pohditaan tarkemmin suhteessa aikaisempaan teoriaan.

7.1 Tutkimustulokset suhteessa aikaisempaan teoriaan

Ensimmäiseksi tutkittiin, mikä teoreettisista AEQ-M-malleista sopisi parhaiten suomalaiseen dataan (katso luku 4.1). Tulokset tukevat hypoteesia 1, jossa esitettiin, että tunne x tilanne -faktorimalli kuvasi parhaiten tunteita suomalaisessa aineistossa. Lisäksi tämän mallin hyvyys oli varsin korkea. Tässä teoreettisessa tunne x tilanne -mallissa matematiikkaan liittyviä tunteita mallinnettiin erillisinä latentteina faktoreina, ja eri tilanteiden vaikutus huomioitiin antamalla samassa tilanteessa olevien muuttujien residuaalien korreloida keskenään. Tämä tutkimustulos tukee aikaisempia tutkimuksia, jotka ovat osoittaneet, että teoreettinen tunne x tilanne -malli kuvaa parhaiten AEQ-M-mittaria (Moreira ym., 2018) ja yleistä AEQ-mittaria (Pekrun ym., 2011). Tämä tutkimus tukee teoreettista näkökulmaa siitä (Pekrun ym., 2011), että akateemisten tunteiden mittaamisessa olisi huomioitava sekä erilliset tunteet että niiden erilaisuus akateemisissa tilanteissa kuten luokassa, oppimisessa ja testissä.

Toiseksi tutkittiin latenttien tunteiden reliabiliteettia ja erotteluväliditeettia. Tutkimustulos tuki hypoteesia siitä, että latenttien tunnemuuttujien reliabiliteetit ovat hyviä. Tämä tulos on linjassa

aikaisempien tutkimuksien kanssa, jotka ovat havainneet korkean reliabiliteetin mittarin latenteille tai havaituille muuttujille (Ahmed ym., 2013; Frenzel ym., 2007b; Holm ym., 2017; Lazarides & Buchholzb, 2019; Moreira ym., 2018; Peixoto ym., 2015; Pekrun ym., 2005).

Tämä tutkimus tuki myös hypoteesia siitä, että mittarin erotteluvaliditeetti on kyseenalainen. Vastaavan ongelman ovat havainneet myös aikaisemmat tutkijat, kun he ovat tutkineet faktorien välisiä korrelaatioita (Moreira ym., 2018; Peixoto ym., 2015; Pekrun ym., 2011). Moreira ym. (2018) havaitsivat korkeita korrelaatioita useiden tunteiden välillä tunne x tilanne -mallissa ja kyseenalaistivat erityisesti vihan ja tylsyyden erotteluvaliditeetin korkean korrelaation perusteella. Tämä tutkimus puolestaan käytti matemaattista kriteeriä tutkimaan erotteluvaliditeettia (katso kaava 3.10.4). Tämä analyysi osoitti, että nautinto ja ylpeys, viha ja ahdistus, ahdistus ja häpeä sekä viha ja tylsyys eivät toteuta erotteluvaliditeetin kriteeriä, eli nämä faktoriparit eivät erotu riittävästi toisistaan. Alhaisen erotteluvaliditeetin taustalla voi olla teoreettinen selitys. Ohjausarvoteorian mukaan tylsyys ja viha liittyvät merkitsemättömiin matematiikan aktiviteetteihin; ahdistus ja häpeä liittyvät matematiikassa epäonnistumiseen; ja nautinto ja ylpeys liittyvät matematiikassa pärjäämiseen ja matematiikan arvostamiseen (Pekrun, 2006). Tämän takia oppilaat voivat kokea näitä tunteita samanaikaisesti ja korrelaatiot ovat korkeita.

Kolmanneksi tämä tutkimus tarkastelee kriittisesti residuaalien välisiä korrelaatioita samassa tilanteessa esiintyvien summamuuttujien välillä teoreettisessa tunne x tilanne -mallissa. Teoreettiset tunne x tilanne -mallit sisältävät näitä korrelaatioita melko paljon ($n > 40$; katso myös Moreira ym., 2018; Pekrun ym., 2011). Itse asiassa tarkempi tarkastelu osoitti, että puolet näistä korrelaatioista ei ollut merkittäviä. Kun nämä ei-merkittävät korrelaatiot poistettiin mallista, mallin hyvyys ei huonontunut merkittävästi. Tämän takia konstruoitu malli (katso kuvio 10) ilman ei-merkittäviä korrelaatioita pitäisi valita kuvaamaan matematiikkaan liittyviä tunteita. Vaikka ohjausarvoteoria (Pekrun, 2006) määrittää tunteiden vahvan yhteyden eri tilanteisiin, tämä tutkimus herättää kysymyksen, ovatko kaikki tunteet samassa tilanteessa kuitenkin merkittävästi yhteydessä. Tulevaisuudessa olisi esitettävä laajempaa analyysia tunteiden tilannesidonnaisuudesta.

Neljänneksi tutkittiin malleja, joissa latentit tunteet, jotka eivät toteuttaneet erotteluvaliditeettia, yhdistettiin samalle faktorille. Tulokset osoittavat, että AEQ-M-mallin hyvyttä ei ole mahdollista parantaa yhdistämällä näitä faktoreita. Tällöin mallin hyvyysindeksien perusteella tunteita pitäisi tarkastella erillisinä tunteina huolimatta siitä, että erotteluvaliditeetti on kyseenalainen. Tämä havainto tukee vahvaa teoreettista ajatusta siitä, että moninaiset tunteet ovat toisiinsa yhteydessä, mutta silti erillisiä (Pekrun, 2006; Pekrun ym., 2011). Tulokset kuitenkin osoittavat, että mittarin erotteluvaliditeetista tulee hyväksyttävä, kun korkeasti korreloivat faktoriparit nautinto ja ylpeys, viha ja tylsyys sekä ahdistus ja häpeä yhdistetään samalle faktorille kolmifaktorisessa mallissa (katso kuvio 11). Tulosten pohjalta olisi pohdittava, onko valittava AEQ-M-malli, jolla on hyvä erotteluvaliditeetti mutta kohtalainen mallin hyvyys, vai AEQ-M-malli, jolla on hyvä mallin hyvyys mutta huono erotteluvaliditeetti. Koska akateemisten tunteiden yhdistämiselle ei ole olemassa teoriapohjaa ja mallin hyvyys huononi merkittävästi faktoreita yhdistäessä, tämä tutkimus tukee näkemystä siitä, että tunteet olisi nähtävä moninaisina ja erillisinä ilmiöinä.

Johtopäätöksenä voidaan todeta, että tämä tutkimus on merkityksellinen, koska se täydentää ohjausarvoteoriaa tunteiden erillisyydestä ja tilannesidonnaisuudesta (Pekrun, 2006). Tämän tutkimustulosten pohjalta voidaan kuitenkin esittää seuraavanlainen kysymys: Olisiko osa tunteista kuitenkin mielekästä yhdistää paremman erotteluvaliditeetin saavuttamiseksi ja liittyvätkö kaikki samassa tilanteessa olevat tunteet aina toisiinsa?

7.2 Tutkimuksen rajoitukset ja kehitysehdotukset

Mittarin kehittäminen ja validointi vaativat jatkotutkimusta. Tässä jatkotutkimuksessa on huomioitava tämän tutkimuksen rajoitukset. Ensinnäkin tämä tutkimus keskittyi suomalaisiin 8. luokan oppilaisiin. Tällöin tutkimusta olisi laajennettava eri ikäryhmille ja eri kulttuurikonteksteihin. Koska tämän tutkimuksen vahvuutena on laaja ja kattava tutkimusotos, on keskeistä jatkossa käyttää ja testata mittaria laajalla otoksella. On myös huomioitava, että tämä uudelleentestaus mahdollistaisi mittarin stabiliteetin arvioimisen (test-retest reliability; Litwin, 1995).

Tämä tutkimus oli poikkileikkaustutkimus. Tällöin pitkittäistutkimuksia tarvitaan selvittämään, onko tunteiden rakenne pysyvä ilmiö vai ajan suhteen muuttuva. Tässä tutkimuksessa tutkittiin tyypillisiä tunteita matematiikassa. Koska mittaria on mahdollista käyttää myös hetkellisten tunteiden tutkimiseen, olisi mittarin teoreettista mallia tutkittava myös suhteessa hetkellisiin tunteisiin. On hyvin mahdollista, että tässä tapauksessa mittarin erotteluvaliditeetti olisi parempi, koska faktorien väliset korrelaatiot voivat olla pienempiä hetkellisille tunteille (Goetz, Preckel, Pekrun & Hall, 2007).

Koska mittari sisälsi varsin vähän eri tilanteisiin liittyviä väittämiä, olisi sitä kehitettävä niin, että eri tilanteisiin lisättäisiin väittämiä. Tämä mahdollistaisi paremmin mittarin tutkimisen suhteessa eri tilanteisiin. Koska mittarissa on jopa 60 väittämää, voi mittarin laajentaminen olla kyseenalaista. Tällöin mittaria voisi olla mielekästä kehittää kolmeksi erilliseksi mittariksi, joissa tunteita mitataan suhteessa eri tilanteissa: luokka, oppiminen ja testi. Tällöin näitä mittareita olisi laajennettava väittämillä niin, että eri mittareista tulee luotettavia. Itse asiassa tätä ratkaisua on käytetty yleisen AEQ-mittarin sovelletun version kohdalla (Peixoto ym., 2015), joka sisältää enemmän väittämiä kuin matematiikkaan suunnattu mittari AEQ-M.

Lisäksi mittarin kehittämisessä olisi huomioitava mittarin erotteluvaliditeetti. Tällöin väittämiä suunniteltaessa olisi tarkkaan mietittävä, kuinka hyvin tunne kuvaa nimenomaan kyseistä tunnetta, kuten esimerkiksi vihaa eikä tylsyyttä. Koska tunteet esiintyvät samanaikaisesti akateemisissa tilanteissa, on hyvin mahdollista, että ei voida rakentaa sellaista tunnemittaria, jossa saavutettaisiin hyvä erotteluvaliditeetti. Tällöin on palattava tutkimuksessa nousevaan kysymykseen siitä, jaetaanko tunteet erillisiksi faktoreiksi, jolloin voidaan saavuttaa tunteiden moninainen esiintyminen oppimisessa. Eiväthän viha ja tylsyys tai nautinto ja ylpeys ole kuitenkaan sama tunne oppilaalle? Lisäksi aikaisempi tutkimus osoittaa, että nämä voimakkaasti korreloivat tunteet ovat kuitenkin eri tavalla yhteydessä matematiikan oppimiseen ja sukupuoleen (Frenzel, Pekrun & Goetz, 2007a; Holm ym., 2017; Pekrun, Lichtenfeld, Marsh, Murayama, & Goetz, 2017). Esimerkiksi oppilaat, joilla on matematiikassa vaikeaa, kokevat nimenomaan voimakkaitten häpeää (Holm ym., 2017). Menetetäänkö tunteista ilmiönä jotakin, jos ne yhdistetään samalla faktorille?

Itse asiassa viimeaikainen tutkimus ehdottaa, että henkilökeskeinen profiilianalyysi ja klusterianalyysi voisivat olla myös hyvä analyysimenetelmä lähestyä akateemisia tunteita (Abd-El-Fattah, 2018; Fernando, Kashima, & Laham, 2014; Ganotice, Datu & King, 2016), koska niiden avulla voidaan tutkia tunteiden moninaista esiintymistä yksilöillä. Tämä menetelmä identifioi yksilöiden joukosta latenteja ryhmiä, joille ovat tyypillistä samanlaiset tunnekokemukset (Howard & Hoffman, 2017).

Ennen kaikkea tämä tutkimus osoittaa, että tulevaisuudessa olisi kehitettävä kriittisesti mittareita ja menetelmiä, jotka mittaavat tunteita ilmiönä mahdollisimman luotettavasti ja monipuolisesti huomioiden niiden erillisyyden, samanaikaisuuden ja tilannesidonnaisuuden.

Lähteet

- Abd-El-Fattah, S. (2018). A validation study of the control-value theory within the domain of mathematics at high school: A latent profile analysis. *Journal of Psychology & Psychotherapy*, 8, 4–15. <http://dx.doi.org/10.4172/2161-0487.1000348>
- Ahmed, W., van der Werf, G., Kuyper, H., & Minnaert, A. (2013). Emotions, self-regulated learning and achievement in mathematics: A growth curve analysis. *Journal of Educational Psychology*, 105, 150–161. <https://doi.org/10.1037/a0030160>
- Bryant, F. B., & Satorra, A. (2012). Principles and practice of scaled difference chi square testing. *Structural Equation Modeling*, 19, 372–398. <http://dx.doi.org/10.1080/10705511.2012.687671>
- Byrne, B. M. (2012). *Structural equation modeling with Mplus: Basic concepts, applications, and programming*. New York: Taylor & Francis Group.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of test. *Psychometrika*, 16, 297–334. <http://dx.doi.org/10.1007/bf02310555>
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B*, 39, 1–38. <http://dx.doi.org/10.2307/2984875>
- Fernando, J. W., Kashima, Y., & Laham, S. M. (2014). Multiple emotions: A person-centered approach to the relationship between intergroup emotion and action orientation. *Emotion*, 14, 722–732. <http://dx.doi.org/10.1037/a0036103>

- Fornell, C., & Larcker, D. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18, 39–50. <https://doi.org/10.1177/002224378101800104>
- Frenzel, A. C., Pekrun, R., & Goetz, T. (2007a). Girls and mathematics—A “hopeless” issue? A control-value approach to gender differences in emotions towards mathematics. *European Journal of Psychology of Education*, 22, 497–514. <http://dx.doi.org/10.1007/BF03173468>
- Frenzel, A. C., Thrash, T. M., Pekrun, R., & Goetz, T. (2007b). Achievement emotions in Germany and China: A cross-cultural validation of the Academic Emotions Questionnaire–Mathematics. *Journal of Cross-Cultural Psychology*, 38, 302–309. <http://dx.doi.org/10.1177/0022022107300276>
- Ganotice, F. A., Datu, J. A. D., & King, R. B. (2016). Which emotional profiles exhibit the best learning outcomes? A person-centered analysis of students’ academic emotions. *School Psychology International*, 37, 498–518. <http://dx.doi.org/10.1177/0143034316660147>
- Geldhof, G., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods*, 19, 72–91. <http://dx.doi.org/10.1037/a0032138>
- Goetz, T., Frenzel, A. C., Pekrun, R., Hall, N. C., & Lüdtke, O. (2007). Between- and within-domain relations of students’ academic emotions. *Journal of Educational Psychology*, 99, 715–733. <http://dx.doi.org/10.1037/0022-0663.99.4.715>

- Goetz, T., Preckel, F., Pekrun, R., & Hall, N. C. (2007). Emotional experiences during test taking: Does cognitive ability make a difference? *Learning and Individual Differences, 17*, 3–16. <https://doi.org/10.1016/j.lindif.2006.12.002>
- Hannula, M. S., & Holm, M. E. (2018). *Oppilaan matematiikkakuva oppimistuloksena ja oppimisen taustatekijänä*. Teoksessa J. Joutsenlahti, H. Silfverberg, & P. Räsänen (Eds.), *Matematiikan opetus ja oppiminen* (pp. 132–154). Jyväskylä: Niilo Mäki Instituutti.
- Holm, M. E., Hannula, M. S., & Björn, P. M. (2017). Mathematics-related emotions among Finnish adolescents across different performance levels. *Educational Psychology, 37*, 205–218. <https://doi.org/10.1080/01443410.2016.1152354>
- Howard, M. C., & Hoffman, M. E. (2017). Variable-centered, person-centered, and person specific approaches: Where theory meets the method. *Organizational Research Methods, 21*, 846–876. <https://doi.org/10.1177/1094428117744021>
- Laaksonen, Seppo. (2013). *Surveyymetodiikka: Aineiston kokoamisesta puhdistamisen kautta analyysiin* (2nd ed.). Retrieved from <https://bookboon.com/fi/surveyymetodiikka-ebook#reviews>
- Lazarides, R., & Buchholzb, J. (2019). Student-perceived teaching quality: How is it related to different achievement emotions in mathematics classrooms? *Learning and Instruction, 61*, 45–59. <https://doi.org/10.1016/j.learninstruc.2019.01.001>
- Lehtonen, R., & Pahkinen, E. (2004). *Practical Methods for Design and Analysis of Complex Surveys* (2nd ed.). Chichester, England: John Wiley & Sons, Ltd.

- Lichtenfeld, S., Pekrun, R., Stupnisky, R., Reiss, K., & Murayama, K. (2012). Measuring students' emotions in the early years: The Achievement Emotions Questionnaire-Elementary School (AEQ-ES). *Learning and Individual Differences*, 22, 190–201. <https://doi.org/10.1016/j.lindif.2011.04.009>
- Litwin, M. 1995. *How to measure survey reliability and validity*. London: Sage Publications Ltd. <https://dx-doi-org.libproxy.helsinki.fi/10.4135/9781483348957>
- Marsh, H. W., Byrne, B. W., & Craven, R. (1992). Overcoming problems in confirmatory factor analyses of MTMM data: The correlated uniqueness model and factorial invariance. *Multivariate Behavioral Research*, 27, 489–507. https://doi.org/10.1207/s15327906mbr2704_1
- Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In search of golden rules: Comment on the hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu & Bentler's (1999) findings. *Structural Equation Modeling*, 11, 320–341. https://doi.org/10.1207/s15328007sem1103_2
- McLachlan, G. J., & Krishnan, T. (2007). *The EM Algorithm and Extensions* (2nd Ed.). New York: John Wiley & Sons.
- Moreira, P., Cunha, D., & Inman, R., A. (2018). Achievement Emotions Questionnaire-Mathematics (AEQ-M) in adolescents: Factorial structure, measurement invariance and convergent validity with personality. *European Journal of Developmental Psychology*, 12, 1–13. <http://dx.doi.org/10.1080/17405629.2018.1548349>

- Musil, C. M., Warner, C. B., Yobas, P. K., & Jones, S. L. A. (2002). A comparison of imputation techniques for handling missing data. *Western Journal of Nursing Research*, 24, 815–829. <https://doi.org/10.1177/019394502762477004>
- Mustonen, Seppo (1995). *Tilastolliset monimuuttujamenetelmät*. Survo Systems. Retrieved from <https://www.survo.fi/mustonen/monim.pdf>
- Muthén, L. K. (2008, February 6). Re: Mplus Discussion [Online forum comment]. Retrieved from <http://www.statmodel.com/discussion/messages/9/352.html>
- Muthén, L. K., & Muthén, B. O. (1998–2013). *Mplus user's guide* (7th ed.). Los Angeles, CA: Authors.
- National Advisory Board on Research Ethics. (2009). *Ethical principles of research in the humanities and social and behavioural sciences and proposals for ethical review.*, Helsinki, Finland: Authors. Retrieved from <http://www.tenk.fi/sites/tenk.fi/files/ethicalprinciples.pdf>
- Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.
- Peixoto, F., Mata, L., Monteiro, V., Sanches, C., & Pekrun, R. (2015). The Achievement Emotions Questionnaire: Validation for pre-adolescent students. *European Journal of Developmental Psychology*, 12, 472–481. <https://doi.org/10.1080/17405629.2015.1040757>
- Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review*, 18, 315–341. <https://doi.org/10.1007/s10648-006-9029-9>

- Pekrun, R., Frenzel, A. C., Goetz, T., & Perry, R. P. (2007). The control-value theory of achievement emotions: An integrative approach to emotions in education. In P. A. Schutz & R. Pekrun (Eds.), *Emotions in education* (pp.13–36). Amsterdam: Academic Press.
- Pekrun, R., Goetz, T., & Frenzel, A. C. (2005). *Academic Emotions Questionnaire – Mathematics (AEQ-M) – User’s manual*. Munich, Germany: University of Munich: Department of Psychology.
- Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., & Perry, R. P. (2011). Measuring emotions in students’ learning and performance: The Achievement Emotions Questionnaire (AEQ). *Contemporary Educational Psychology*, 36, 36–48. <https://doi.org/10.1016/j.cedpsych.2010.10.002>
- Pekrun, R., Lichtenfeld, S., Marsh, H. W., Murayama, K., & Goetz, T. (2017). Achievement emotions and academic performance: Longitudinal models of reciprocal effects. *Child Development*, 88, 1653–1670. <https://doi.org/10.1111/cdev.12704>
- Rencher, A. C., & Christensen, W. F. (2012). *Methods of multivariate analysis* (3rd ed.). New Jersey: John Wiley & Sons, Inc., Hoboken. <http://dx.doi.org/10.1002/9781118391686>
- Rubin, L. H., Witkiewitz, K., St. Andre, J., & Reilly, S. (2007). Methods for handling missing data in the behavioral neurosciences: Don’t throw the baby rat out with the bath water. *Journal of Undergraduate Neuroscience Education*, 5, A71–A77. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3592650/>

- Satorra, A. & Bentler, P.M. (2010). Ensuring positiveness of the scaled difference chi-square test statistic. *Psychometrika* 75, 243–248. <https://doi.org/10.1007/s11336-009-9135-y>
- Scheffer, J. (2002). Dealing with missing data. *Research Letters in the Information and Mathematical Sciences*, 3, 153–160. Retrieved from http://exec-ed.ac.nz/massey/fms/Colleges/College%20of%20Sciences/IIMS/RLIMS/Volume03/Dealing_with_Missing_Data.pdf
- Scherer, K. (2005). What are emotions? And how can they be measured? *Social Science Information*, 44, 695–729. <https://doi.org/10.1177/0539018405058216>
- Schutz, P. A., & Lanehart, S. L. (2002). Emotions in education. *Educational Psychologist* 37, 67–68. https://doi.org/10.1207/S15326985EP3702_1
- Vehkalahti, K. (2014). *Kysely tutkimuksen mittarit ja menetelmät*. Oy Finn Lectura Ab.
- Vehkalahti, K., & Everitt, B. S. (2019). *Multivariate Analysis for the Behavioral Sciences* (2nd Ed.). Florida: CRC Press, Taylor & Francis Group.
- Weiner, B. (1985). Attributional theory of achievement motivation and emotion. *Psychology Review*, 92, 548–573. <http://dx.doi.org/10.1037/0033-295X.92.4.548>
- Weiner, B. (2014). The attribution approach to emotion and motivation: History, hypotheses, home runs, headaches/heartaches. *Emotion Review*, 6, 353–361. <http://dx.doi.org/10.1177/1754073914534502>

Zait, A., & Berteau, P. E. (2011). Methods for testing discriminant validity. *Management & Marketing* 9(2), 217–224. Retrieved from <https://doaj.org/article/e01ef0e719424d478aac588d2cc925bf>

Zeidner, M. (2007). *Test anxiety in educational contexts: What I have learned so far*. In P. A. Schutz & R. Pekrun (Eds.), *Emotion in education* (pp. 165–184). San Diego, CA: Academic Press. <http://dx.doi.org/10.1016/B978-012372545-5/50011-3>